Low-rank and sparse reconstruction in dynamic magnetic resonance imaging via proximal splitting methods

Benjamin R. Trémoulhéac

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London

Centre for Medical Image Computing
Department of Computer Science
Department of Medical Physics and Biomedical Engineering

December 2014

Declaration

I, Benjamin R. Trémoulhéac, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Benjamin R. Trémoulhéac
December 2014

Abstract

Dynamic magnetic resonance imaging (MRI) consists of collecting multiple MR images in time, resulting in a spatio-temporal signal. However, MRI intrinsically suffers from long acquisition times due to various constraints. This limits the full potential of dynamic MR imaging, such as obtaining high spatial and temporal resolutions which are crucial to observe dynamic phenomena.

This dissertation addresses the problem of the reconstruction of dynamic MR images from a limited amount of samples arising from a nuclear magnetic resonance experiment. The term limited can be explained by the approach taken in this thesis to speed up scan time, which is based on violating the Nyquist criterion by skipping measurements that would be normally acquired in a standard MRI procedure. The resulting problem can be classified in the general framework of linear ill-posed inverse problems. This thesis shows how low-dimensional signal models, specifically low-rank and sparsity, can help in the reconstruction of dynamic images from partial measurements. The use of these models are justified by significant developments in signal recovery techniques from partial data that have emerged in recent years in signal processing.

The major contributions of this thesis are the development and characterisation of fast and efficient computational tools using convex low-rank and sparse constraints via proximal gradient methods, the development and characterisation of a novel joint reconstruction–separation method via the low-rank plus sparse matrix decomposition technique, and the development and characterisation of low-rank based recovery methods in the context of dynamic parallel MRI. Finally, an additional contribution of this thesis is to formulate the various MR image reconstruction problems in the context of convex optimisation to develop algorithms based on proximal splitting methods.

Acknowledgement

I would like to thank my supervisors Simon Arridge and David Atkinson for their support and guidance. These years were a great opportunity to introduce me into academic research, in such a fascinating city that is London. I would like to acknowledge financial support from EPSRC. I'd like to thank both Dominique and Ron for making things easier in the lab. I would like to thank friends and colleagues at UCL, in particular Nikos, Uran, Luis, Elwin, Bjoern. My thoughts are also with Philip Batchelor.

On a more personal note, I am grateful to have met many lovely people in London. A special thank to David and Andrew for their support and friendship throughout these years. I also would like to thank the following Londoners: Alex, Aline, Antoine, Julie, Marin, Nathalie, Olivier, Romain, Yannick. I'd like to greet the former *guerriers*, Étienne, Mathieu, Rémi, Flo, Julien and the rest of the team. Thanks to Dr. Alex for his advices, and thanks to my brother and his family. Finally, thanks to Diane for those moments you have shared with me, et évidemment, merci à vous papa et maman.

Contents

List of figures

List of tables

List of algorithms

List of acronyms and abbreviations

2D	Two-dimensional
ADMM	Alternating direction method of multipliers
AL	Augmented Lagrangian
ALM	Augmented Lagrangian method
CS	Compressed sensing
CT	Computed tomography
dB	Decibel
DCE	Dynamic contrast-enhanced
DFT	Discrete Fourier transform
FCS	Fast composite splitting
FISTA	Fast iterative shrinkage-thresholding algorithm
FFT	Fast Fourier transform
FOV	Field of view
FPG	Fast proximal gradient
GFBS	Generalised forward-backward splitting
i.i.d.	Independent and identically distributed
IDFT	Inverse discrete Fourier transform
MM	Method of multipliers
MR	Magnetic resonance
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
NMSE	Normalised mean square error
PCA	Principal component analysis
PMRI	Parallel magnetic resonance imaging
PVD	Polynomial variable density
RF	Radio frequency
RIP	Restricted isometry property
RPCA	Robust principal component analysis
s.t.	subject to

SNR	Signal-to-noise ratio
SSoS	Square root of sum of squares
SVD	Singular value decomposition
TV	Total variation

List of symbols

Common symbols

i		Imaginary unit
$\Re$, $\Im$		Real part, imaginary part
$.^\top$, $.^\mathrm{H}$		Transpose, conjugate transpose
$\bar{.}$		Complex conjugate
$.^*$		Adjoint
$.^\star$		Optimal point
$\hat{.}$		Estimation
$\mathcal{N}$		Normal distribution
∂		Partial derivative
$\alpha, \beta, \mu, \lambda$	$\in \mathbb{R}^+$	Regularisation parameters
ρ	$\in \mathbb{R}_*^+$	Step size
L	$\in \mathbb{R}_*^+$	Lipschitz constant
δ_K, δ_R	$\in \mathbb{R}^+$	Isometry constants for K-sparse signal and R-rank matrix
δ	$\in \mathbb{R}^+$	Penalty parameters for augmented Lagrangian
$\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$	$\in \mathbb{R}^3$	Unit vectors in the Cartesian coordinate system
D	$\in \{1, 2, 3\}$	Signal dimension for continuous data
$\mathbf{r}$	$\in \mathbb{R}^D$	Position coordinates for continuous data
$\mathbf{k}$	$\in \mathbb{R}^D$	**k**-space coordinates for continuous data

Functions

S, S_j	$\in \mathbb{C}$	**k**-space function, **k**-space function for coil j
Γ_j	$\in \mathbb{C}$	Sensitivity map for coil j
I	$\in \mathbb{C}$	Image function
E	$\in \mathbb{C}$	Encoding function
N	$\in \mathbb{C}$	Noise function
F	$\in \mathbb{R}$	Objective function
$\mathcal{L}$, $\mathcal{L}^A$	$\in \mathbb{R}$	Lagrangian function, augmented Lagrangian function
$\mathbf{1}_C$	$\in \mathbb{R} \cup \{+\infty\}$	Characteristic function of C [Eq. (4.2)]

Discrete data

$N, M, N_1, N_2, \ldots$	$\in \mathbb{N}$	Vector or matrix dimension
N_x	$\in \mathbb{N}$	Number of pixels for x-coordinate
N_y	$\in \mathbb{N}$	Number of pixels for y-coordinate
N_t	$\in \mathbb{N}$	Number of temporal frames
N_c	$\in \mathbb{N}$	Number of coils
$\mathbf{y}$	$\in \mathbb{C}^M$	**k**-space measurement vector
$\mathbf{n}$	$\in \mathbb{C}^M$	Noise vector
$\mathbf{I}$	$\in \mathbb{R}^{N \times N}$	Identity matrix
$\mathbf{A}$	$\in \mathbb{R}^{M \times N}$	Sampling matrix
$\mathbf{\Psi}$	$\in \mathbb{C}^{N \times N}$	Sparsifying transform matrix
$\mathbf{F}$	$\in \mathbb{C}^{N \times N}$	Fourier transform matrix
$\mathbf{E}$	$\in \mathbb{C}^{M \times N}$	MRI encoding matrix
$\boldsymbol{\mathcal{X}}$	$\in \mathbb{C}^{N_1 \times N_2 \times \ldots \times N_Q}$	Q-order tensor

Operators

$\mathcal{A}$		Sampling operator
Ψ		Sparsifying transform operator
$\mathcal{F}$		Fourier transform operator
$\mathcal{F}_t$		Temporal Fourier transform operator
$\mathcal{E}$		MRI encoding operator
$\mathcal{E}_\Gamma$		Parallel MRI encoding operator
prox_g		Proximal operator of g
∇		Gradient operator
$\mathcal{S}$	$\mathbb{C}^N \to \mathbb{C}^N$	Soft thresholding operator [Eq. (4.11)]
$\mathcal{S}^p$	$\mathbb{C}^N \to \mathbb{C}^N$	Generalised soft thresholding operator [Eq. (5.14)]
$\mathcal{H}$	$\mathbb{C}^N \to \mathbb{C}^N$	Hard thresholding operator [Eq. (4.52)]
$\text{SVT}^{\mathcal{S}}$	$\mathbb{C}^{M \times N} \to \mathbb{C}^{M \times N}$	Singular values soft thresholding operator [Eq. (4.14)]
$\text{SVT}^{\mathcal{S}^p}$	$\mathbb{C}^{M \times N} \to \mathbb{C}^{M \times N}$	Singular values generalised soft thresholding operator [Eq. (5.15)]
$\text{SVT}^{\mathcal{H}}$	$\mathbb{C}^{M \times N} \to \mathbb{C}^{M \times N}$	Singular values hard thresholding operator [Eq. (4.53)]

Chapter 1

Introduction

Contents

1.1 Motivations

Seeing inside the living body without the need for exploratory surgery has always been a challenge in human history. This has been possible since X-rays were discovered in 1895 by Röntgen, who produced the first picture showing bone structures of his wife's hand. From that time, a number of medical imaging techniques have been developed, such as positron emission tomography, X-ray computed tomography (CT), nuclear magnetic resonance imaging or ultrasound.

Nuclear magnetic resonance imaging, or simply magnetic resonance imaging (MRI), is a medical imaging technique that is based primarily upon the sensitivity to the presence and properties of water. MRI uses magnetic fields and radio electromagnetic waves to detect tiny changes in the magnetism of the nucleus of the hydrogen atom which is found in abundance in the human body. MRI is a valuable diagnostic tool that is extensively used in radiology to examine the anatomy and physiology of the body. It is generally regarded as a safe procedure because it does not involve ionising radiation in contrast to X-ray CT for example. MRI can also produce different types of images without any mechanical modification to the MRI scanner, such as different image contrasts or images of the subject in various orientations and positions.

In this thesis, we are interested in dynamic MRI, a technique that consists of collecting MR images in time and thus generating a spatio-temporal signal. Dynamic

MRI is used in multiple clinical applications, such as cardiovascular (or cardiac) MRI, dynamic contrast-enhanced (DCE) MRI or functional MRI (fMRI). Cardiac MR imaging is used to capture and study quantitative assessment of the heart. A primary application is cine cardiac imaging, which assesses both structure and function of the beating heart. Other applications include for example myocardial perfusion imaging to detect coronary artery disease or phase contrast imaging to quantify blood flow velocity through the heart. Regarding DCE MRI, it is used to assess the passage and distribution of a contrast agent through organs and may possibly be used as a source of biomarkers in oncology. An important example is breast cancer imaging based on gadolinium contrast agent. In neuroimaging studies, functional MRI is used to localise brain activities by detecting blood-oxygen-level-dependent signals.

However, one of the fundamental limiting factor of MRI is its serial acquisition procedure that is inherently slow due to various constraints. These include intrinsic nuclear relaxation times that generate the signal to acquire, how fast gradient magnetic fields can be switched on and off without causing peripheral nerve stimulations, how fast the oscillating radio frequency magnetic field can be turned on and off to prevent tissue heating (specific absorption rate), and signal-to-noise ratio constraints. Although it is possible to image static objects such as the brain with a slow acquisition procedure, it is much more problematic to collect images of moving structures such as the beating heart or in which contrast changes over time, as in dynamic MR imaging.

This thesis is motivated by the benefits of reducing the acquisition time of dynamic MRI while maintaining the image quality. From the patient perspective, it increases comfort, facilitates scans for problematic subjects such as the very young, old or ill, and also limits patient exposure to magnetic fields and acoustic noise. From an imaging aspect, obtaining faster scan times can benefit every dynamic MRI applications mentioned previously, since higher temporal resolution better characterises dynamic processes, or trading the time saving for higher spatial resolution provides greater anatomical details. Minimisation of the patient's time in the scanner also decreases the chances of motion artefacts such as blurring in the resulting images.

1.2 Problem statement

The main problem addressed in this thesis is the reconstruction of spatio-temporal magnetic resonance images from a limited amount of samples acquired in the Fourier domain, known as $(\mathbf{k}, t)$-space in MRI. The term *limited* can be explained by the approach taken in this thesis to speed up scan time, which is based on violating the Nyquist criterion by skipping measurements that would be normally acquired in a standard MRI procedure.

Due to the physics of MRI, the spatio-temporal MRI signal $S(\mathbf{k}, t)$ can be modelled mathematically through the Fourier integral as

$$S(\mathbf{k}, t) = \int I(\mathbf{r}, t) e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})} d\mathbf{r} + N(\mathbf{k}, t), \tag{1.1}$$

where $I(\mathbf{r}, t)$ and $N(\mathbf{k}, t)$ are respectively the spatio-temporal image and noise functions, and $\mathbf{k}$ and $\mathbf{r}$ are respectively the $\mathbf{k}$-space and position coordinates. The image reconstruction task can be expressed as: given a finite set of sub-Nyquist Fourier measurements from Eq. (1.1), find the best discrete approximation of the spatio-temporal function $I(\mathbf{r}, t)$. This represents a typical example of a *linear ill-posed inverse problem*.

1.3 Thesis objectives and contributions

This thesis shows how incorporating prior information based on low-dimensional signal models can help in the reconstruction of spatio-temporal images from a limited amount of Fourier samples arising from a nuclear magnetic resonance experiment.

More specifically, this thesis investigates the use of *low-rank and sparse* signal models through *proximal splitting methods* to help in the reconstruction of *dynamic MR images* from partial data. Low-rank and sparsity models have contributed to significant developments in signal recovery techniques in recent years in the fields of signal processing and applied mathematics. These models can be classified as low-dimension or low-complexity as they are related to the principle of parsimony, also known as Occam's razor, which states that the simplest among competing hypotheses should be preferred. The use of proximal splitting methods in this thesis are encouraged by the requirement of reconstruction algorithms (i) to handle nonsmooth penalties due to low-rank and sparse constraints, and (ii) to tackle relatively large-scale problems due to spatio-temporal MR signals lying in high-dimensional spaces. This thesis aims to provide an adequate trade-off between the theoretical concepts of signal recovery and the practical aspects of the MRI reconstruction problem from limited Fourier measurements. In short, the major contributions of this thesis are

- the development and characterisation of computational methods for low-rank and sparsity constrained problems based on fast proximal gradient methods,
- the development and characterisation of a joint reconstruction–separation model that goes beyond traditional reconstruction methods,
- the development and characterisation of low-rank based recovery methods in combination with sub-Nyquist dynamic parallel MR imaging,

- the formulation of the various MR image reconstruction problems in the context of convex optimisation and proximal splitting methods.

These contributions are further explained in section 1.4 which presents the outline of this manuscript. Some of the work presented in this thesis have been previously published in journal article:

- B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge. Dynamic MR image reconstruction–separation from undersampled (k,t)-space via low-rank plus sparse prior. *IEEE Transactions on Medical Imaging*, 33(8):1689–1701, 2014.

And in international conference papers:

- B. Trémoulhéac, D. Atkinson, and S. R. Arridge. Low-rank and (x-f)-space sparsity via fast composite splitting for accelerated dynamic MR imaging. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 649–652, Beijing, 2014.

- B. Trémoulhéac, D. Atkinson, and S. R. Arridge. Fast dynamic MRI via nuclear norm minimization and accelerated proximal gradient. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 322–325, San Francisco, 2013.

- B. Trémoulhéac, D. Atkinson, and S. R. Arridge. Motion and contrast enhancement separation model reconstruction from partial measurements in dynamic MRI. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on Sparsity Techniques in Medical Imaging*, Nice, 2012.

Work during this thesis has also resulted in other publications as co-author in international conferences, although not reported in this manuscript.

1.4 Outline

The necessary background on magnetic resonance imaging is introduced in chapter 2. We take care of describing the MRI signal path from the most simplest raw nuclear magnetic resonance signal (nuclear magnets) to arrive at the Fourier integral transform. Other important topics covered include image reconstruction, noise issues and dynamic MR imaging.

Chapter 3 describes linear inverse problems, a fundamental topic not only in medical imaging but in many other areas of science and engineering. Most of this chapter is concerned with signal recovery from partial data. We emphasise the

description of signal recovery methods through the use of low-dimensional signal models that have gained much interest in the past decade. We then discuss state of the art recovery techniques for the specific problem of dynamic MRI reconstruction.

Chapter 4 introduces succinctly proximal splitting methods, a general optimisation framework that provides efficient and flexible algorithms to minimise certain types of convex problems. This framework forms the basis to solve the different formulated image reconstruction problems in this thesis.

Chapter 5 develops and characterises efficient computational tools based on proximal gradient for MR image reconstruction that exploits sparse and low-rank structures. Characterisation of these methods are shown with multiple realistic datasets and various comparisons with other state of the art methods.

Chapter 6 proposes a joint reconstruction–separation model that goes beyond traditional reconstruction methods from partial observations. The model is based on the low-rank plus sparse matrix decomposition to both regularise and intrinsically separate reconstructed dynamic data. The proposed technique provides a competitive reconstruction method, as well as the ability to separate clinically-relevant data in the context of dynamic contrast enhanced MRI.

In chapter 7, we explore low-rank based recovery approaches in combination with dynamic parallel imaging. Parallel imaging is the most widely used technique to accelerate imaging scan in clinical practice. We show how low-rank based signal recovery techniques can be combined with dynamic parallel imaging in various ways to enable further improvement in image reconstruction.

We conclude this thesis in chapter 8 by providing a summary of contributions, the current limitations and perspectives of this thesis.

Chapter 2

Principles of magnetic resonance imaging

Contents

2.1 Introduction

In this chapter, fundamental concepts of magnetic resonance imaging (MRI) are presented. Due to the complicated nature of an MRI system however, we only present an overview of the physics of MRI. The reader is referred to Refs. [1–3] for a more in-depth analysis.

This chapter is organised as follows. We discuss the nuclear magnetic resonance (NMR) phenomenon in section 2.2. The concepts of signal detection, spatial localisation and Fourier encoding are then described respectively in sections 2.3, 2.4 and

2.5. Note that from section 2.2 to 2.5, we describe consecutively the various forms and transitions of the MRI signal as

$$\boldsymbol{\mu} \longrightarrow \mathbf{M} \longrightarrow M_+(t) \longrightarrow V(t) \longrightarrow S(t) \longrightarrow S(\mathbf{k}),$$

where $\boldsymbol{\mu}$ represents the nuclear magnetic moment, $\mathbf{M}$ an ensemble of spins, $M_+(t)$ the transverse magnetisation, $V(t)$ the voltage signal, $S(t)$ the nuclear magnetic resonance signal, to finally $S(\mathbf{k})$ the $\mathbf{k}$-space signal. The transition from continuous to discrete data is then explained in the image reconstruction section 2.6. Noise in MRI is discussed in section 2.7, dynamic imaging in section 2.8 and we conclude this chapter in section 2.9.

2.2 Nuclear magnetic resonance phenomenon

MRI relies on the nuclear magnetic resonance (NMR) phenomenon which involves the interaction of atomic nuclei in a magnetic field. It was first observed independently by Bloch and Purcell in 1946 for which they shared the Nobel Prize in Physics in 1952. To understand the NMR phenomenon, we should in principle start at the nuclear level using laws of quantum theory because the behaviour of atomic and subatomic particles can only be accurately described in this framework. However, the phenomenon can be explained more simplistically using the theory of classical (Newtonian) mechanics. It is possible to do so because MRI deals with the collective behaviour of a large ensemble of particles.

Atoms consist of nuclei (protons and neutrons) surrounded by their orbiting electrons. A nucleus has a finite radius, mass and a net electric charge. Some nuclei, depending on their atomic weights or numbers, possess an angular momentum $\mathbf{J}$ often called *spin*, such as the nucleus of the hydrogen atom present in water. This angular momentum combined with the electric charge of the nucleus induces a magnetic field known as the *nuclear magnetic moment* $\boldsymbol{\mu}$. The nuclear magnetic moment can be expressed as

$$\boldsymbol{\mu} = \mu_x\hat{\mathbf{i}} + \mu_y\hat{\mathbf{j}} + \mu_z\hat{\mathbf{k}}, \tag{2.1}$$

where $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$ denote the unit directional vectors in the standard Cartesian coordinate system and (μ_x, μ_y, μ_z) the scalar components of $\boldsymbol{\mu}$. The angular momentum and the magnetic moment are related via $\boldsymbol{\mu} = \gamma\mathbf{J}$, where γ is a nucleus dependent constant known as the *gyromagnetic ratio*. Nuclei with nonzero $\boldsymbol{\mu}$ are then regarded as microscopic magnets. Nuclear spin can be visualised as a physical rotation of the nucleus about its own axis, although it is a property characterised by quantum mechanics.

If we consider an ensemble of spins, nuclear magnets can be added together. The

total magnetic field $\mathbf{M}$ is referred to as the *bulk magnetisation vector*,

$$\mathbf{M} = \sum_n \boldsymbol{\mu}_n = M_x\hat{\mathbf{i}} + M_y\hat{\mathbf{j}} + M_z\hat{\mathbf{k}}, \tag{2.2}$$

where the n^{th} nucleus has magnetic moment $\boldsymbol{\mu}_n$. Magnitudes of each magnetic moment $|\boldsymbol{\mu}_n|$ are known under any conditions, but their directions are completely random due to thermal movements. This results in no net magnetisation, $\mathbf{M} = \mathbf{0}$.

However, in the presence of an external magnetic field applied in the z-direction $\mathbf{B}_0 = B_0\hat{\mathbf{k}}$, nuclear magnetic moments for the hydrogen are restricted to two possible orientations: parallel to $\mathbf{B}_0$ (alignment, spin up) or antiparallel to $\mathbf{B}_0$ (anti-alignment, spin down). On average, more nuclear spins align with the magnetic field than against, which results in a net bulk thermal equilibrium magnetisation in the z-direction

$$\mathbf{M} = M_z^0\hat{\mathbf{k}} \tag{2.3}$$

where M_z^0 represents the thermal equilibrium value for $\mathbf{M}$ in the presence of $\mathbf{B}_0$.

Apart from forcing them to align, each magnetic moment experiences a torque from $\mathbf{B}_0$ causing it to precess about the $\hat{\mathbf{k}}$ axis. This phenomenon is called *nuclear precession* and can be physically interpreted as the wobbling of a spinning-top about the gravitational axis, or the precession of a spinning toy gyroscope. The precession frequency of $\boldsymbol{\mu}$ experiencing a $\mathbf{B}_0$ field is given by the Larmor equation

$$\omega_0 = \gamma|\mathbf{B}_0| = \gamma B_0. \tag{2.4}$$

We consider now an oscillating magnetic field $\mathbf{B}_1$ in addition to the static $\mathbf{B}_0$ field. (The strength of $\mathbf{B}_1$ is much weaker than $\mathbf{B}_0$.) The application of this oscillating magnetic field causes $\mathbf{M}$ to tip away from the $\mathbf{B}_0$ field at a specific angle, known as the *flip angle*. The $\mathbf{B}_1$ field is often referred to as *radio frequency (RF) pulse* because it oscillates in the radio frequency range and it is turned on only for a few milliseconds. The resonance condition in magnetic *resonance* imaging is that $\mathbf{B}_1$ should rotate in the same manner as the precessing spins, i.e. rotating at the Larmor frequency ω_0 around the z-direction.

In the presence of the external magnetic field $\mathbf{B}_1(t)$, the time-dependent behaviour of the bulk magnetisation vector $\mathbf{M}$ can be described according to the Bloch equation

$$\frac{\partial \mathbf{M}}{\partial t} = \gamma\mathbf{M} \times \mathbf{B} - \frac{1}{T_2}(M_x\hat{\mathbf{i}} + M_y\hat{\mathbf{j}}) + \frac{1}{T_1}(M_z^0 - M_z)\hat{\mathbf{k}}, \tag{2.5}$$

where $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1(t)$ is the total magnetic field experienced by the nuclei; T_1 and T_2 are decay constants called *relaxation times* that characterise the relaxation process after the spin system has been perturbed (they depend on the different tissues such as white/grey matter or fat); and M_z^0 is the thermal equilibrium value for $\mathbf{M}$ in the

presence of $\mathbf{B}_0$ only.

According to the laws of thermodynamics, and provided the RF pulse is turned off and sufficient time is given, the spin system relaxes back towards its original equilibrium state. This phenomenon is called *relaxation* and is characterised by a precession of $\mathbf{M}$ about the $\mathbf{B}_0$ field, a recovery of the longitudinal magnetisation M_z (longitudinal relaxation) and the destruction of the transverse magnetisation M_+ (transverse relaxation). From the solutions to the Bloch equation, the time evolution for the longitudinal and transverse magnetisations are

$$M_z(t) = M_z^0(1 - e^{-t/T_1}) + M_z(0)e^{-t/T_1}, \tag{2.6a}$$

$$M_+(t) = M_+(0)e^{-t/T_2}e^{-i\omega_0 t}, \tag{2.6b}$$

where $M_z(0)$ and $M_+(0)$ are respectively the magnetisation along the z-direction and on the transverse plane after an RF pulse. The transverse relaxation M_+ uses the complex representation

$$M_+(t) = M_x(t) + iM_y(t), \tag{2.7}$$

with

$$\begin{aligned} M_x(t) &= e^{-t/T_2}\Big(M_x(0)\cos(\omega_0 t) + M_y(0)\sin(\omega_0 t)\Big), \\ M_y(t) &= e^{-t/T_2}\Big(M_y(0)\cos(\omega_0 t) - M_x(0)\sin(\omega_0 t)\Big). \end{aligned} \tag{2.8}$$

The complex formulation is useful because the main activity of the spin in a static magnetic field is a rotation in the 2D transverse plane.

2.3 Signal detection

Signal detection relies on the Faraday law of electromagnetic induction. This law states that a time-varying magnetic flux $\Phi(t)$ (in Webers) through an electric circuit produces an electromagnetic force $\mathcal{E}(t)$ (in Volts). More specifically, the electromagnetic force $\mathcal{E}(t)$ induced in the circuit is equal to the negative of the time rate of change of the magnetic flux,

$$\mathcal{E}(t) = -\frac{\partial \Phi(t)}{\partial t}. \tag{2.9}$$

According to the principle of reciprocity, the magnetic flux in MRI can be expressed as

$$\Phi(t) = \int \mathbf{B}^r(\mathbf{r}) \cdot \mathbf{M}(\mathbf{r}, t) d\mathbf{r}, \tag{2.10}$$

where $\mathbf{B}^r(\mathbf{r}) = B_x^r(\mathbf{r})\hat{\mathbf{i}} + B_y^r(\mathbf{r})\hat{\mathbf{j}} + B_z^r(\mathbf{r})\hat{\mathbf{k}}$ is the magnetic field received in the coil and $\mathbf{M}(\mathbf{r}, t)$ is the bulk magnetisation vector, both at position $\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$.

Hence, the electromagnetic force is given by

$$\mathcal{E}(t) = -\frac{\partial}{\partial t}\int \mathbf{B}^r(\mathbf{r}) \cdot \mathbf{M}(\mathbf{r},t)d\mathbf{r}. \tag{2.11}$$

The voltage signal $V(t)$ induced in the MRI system coil is proportional to Eq. (2.11) depending on the characteristics of the measurement system. Explicitly, it can be written as

$$V(t) \propto -\frac{\partial}{\partial t}\int \left[B_x^r(\mathbf{r})M_x(\mathbf{r},t) + B_y^r(\mathbf{r})M_y(\mathbf{r},t) + B_z^r(\mathbf{r})M_z(\mathbf{r},t)\right] d\mathbf{r}. \tag{2.12}$$

To develop expression (2.12), first note that the previous static-field solutions (2.6a) and (2.6b) hold for each $\mathbf{r}$,

$$M_z(\mathbf{r},t) = M_z^0(1 - e^{-t/T_1(\mathbf{r})}) + M_z(\mathbf{r},0)e^{-t/T_1(\mathbf{r})}, \tag{2.13a}$$
$$M_+(\mathbf{r},t) = M_+(\mathbf{r},0)e^{-t/T_2(\mathbf{r})}e^{-i\omega_0 t} = |M_+(\mathbf{r},0)|e^{i\phi_0(\mathbf{r})}e^{-t/T_2(\mathbf{r})}e^{-i\omega_0 t}, \tag{2.13b}$$

where the phase ϕ_0 and magnitude $|M_+(\mathbf{r},0)|$ are determined by the initial RF pulse conditions in Eq. (2.13b). We now address the computation of the time derivatives in Eq. (2.12). For $\partial_t M_z$, we have

$$\begin{aligned}\frac{\partial M_z(\mathbf{r},t)}{\partial t} &= \frac{\partial}{\partial t}\Big(M_z^0(1 - e^{-t/T_1(\mathbf{r})}) + M_z(\mathbf{r},0)e^{-t/T_1(\mathbf{r})}\Big) \\ &= e^{-t/T_1(\mathbf{r})}\Big(\frac{M_z^0}{T_1(\mathbf{r})} - \frac{M_z(\mathbf{r},0)}{T_1(\mathbf{r})}\Big).\end{aligned} \tag{2.14}$$

For $\partial_t M_x$ and $\partial_t M_y$, it is useful to employ the complex representation (2.7) with the fact that $M_x = \Re\{M_+\}$ and $M_y = \Im\{M_+\}$,

$$\begin{aligned}\frac{\partial M_x(\mathbf{r},t)}{\partial t} = \frac{\partial \Re\{M_+(\mathbf{r},t)\}}{\partial t} &= \Re\Big\{\frac{\partial}{\partial t}M_+(\mathbf{r},0)e^{-t/T_2(\mathbf{r})}e^{-i\omega_0 t}\Big\} \\ &= -\Re\Big\{M_+(\mathbf{r},0)\Big(\frac{1}{T_2(\mathbf{r})} + i\omega_0\Big)e^{-t/T_2(\mathbf{r})}e^{-i\omega_0 t}\Big\},\end{aligned} \tag{2.15}$$

$$\frac{\partial M_y(\mathbf{r},t)}{\partial t} = \frac{\partial \Im\{M_+(\mathbf{r},t)\}}{\partial t} = -\Im\Big\{M_+(\mathbf{r},0)\Big(\frac{1}{T_2(\mathbf{r})} + i\omega_0\Big)e^{-t/T_2(\mathbf{r})}e^{-i\omega_0 t}\Big\}. \tag{2.16}$$

However, since the Larmor frequency ω_0 is generally orders of magnitude larger than typical values of $1/T_1$ and $1/T_2$, the time derivatives $\partial_t M_x$ and $\partial_t M_y$ can be approximated to

$$\partial_t M_x(\mathbf{r},t) \approx -\omega_0 e^{-t/T_2(\mathbf{r})}\Re\{iM_+(\mathbf{r},0)e^{-i\omega_0 t}\}, \tag{2.17a}$$
$$\partial_t M_y(\mathbf{r},t) \approx -\omega_0 e^{-t/T_2(\mathbf{r})}\Im\{iM_+(\mathbf{r},0)e^{-i\omega_0 t}\}, \tag{2.17b}$$

and expression (2.12) can be simplified to

$$V(t) \propto -\int \left[B_x^r(\mathbf{r})\partial_t M_x(\mathbf{r},t) + B_y^r(\mathbf{r})\partial_t M_y(\mathbf{r},t)\right] d\mathbf{r}. \tag{2.18}$$

The above expression shows that the dominant signal induced in the receiver coil is due to the transverse magnetisation and not the longitudinal magnetisation. Substituting the approximated derivatives (2.17a) and (2.17b) into (2.18) yields

$$V(t) \propto \omega_0 \int e^{-t/T_2(\mathbf{r})} \left[B_x^r(\mathbf{r})\Re\{iM_+(\mathbf{r},0)e^{-i\omega_0 t}\} + B_y^r(\mathbf{r})\Im\{iM_+(\mathbf{r},0)e^{-i\omega_0 t}\right] d\mathbf{r}. \tag{2.19}$$

Using the fact that $M_+(\mathbf{r},0) = |M_+(\mathbf{r},0)|e^{i\phi_0(\mathbf{r})}$ from Eq. (2.13b), the above expression reads

$$V(t) \propto \omega_0 \int e^{-t/T_2(\mathbf{r})}|M_+(\mathbf{r},0)| \left[B_x^r(\mathbf{r})\sin(\omega_0 t - \phi_0(\mathbf{r})) + B_y^r(\mathbf{r})\cos(\omega_0 t - \phi_0(\mathbf{r}))\right] d\mathbf{r}. \tag{2.20}$$

This expression can be further simplified to

$$V(t) \propto \omega_0 \int |B_+^r(\mathbf{r})||M_+(\mathbf{r},0)|e^{-t/T_2(\mathbf{r})}\sin(\omega_0 t + \phi_r(\mathbf{r}) - \phi_0(\mathbf{r}))d\mathbf{r} \tag{2.21}$$

by defining the magnetic field received components as

$$B_x^r(\mathbf{r}) = |B_+^r(\mathbf{r})|\cos(\phi_r(\mathbf{r})), \tag{2.22a}$$

$$B_y^r(\mathbf{r}) = |B_+^r(\mathbf{r})|\sin(\phi_r(\mathbf{r})), \tag{2.22b}$$

where $\phi_r(\mathbf{r})$ is the reception phase angle.

At this stage, $V(t)$ is a high-frequency signal because the transverse magnetisation vector precesses at the Larmor frequency ω_0, which is about 42.5 MHz per Tesla for hydrogen protons. To avoid potential problems in later stages, this signal is moved to a low-frequency band in order to be free of the high Larmor oscillation. In practice, it consists of multiplying the time signal $V(t)$ by a complex exponential with frequency $\Omega = \omega_0 + \delta\omega$ where $\delta\omega$ is a small offset frequency, or equivalently multiplying $V(t)$ separately with a cosine and sine signals at the frequency Ω. Each separate signal then results in two components, one close to the offset frequency and the other nearly twice the Larmor frequency. Both signals are then low-pass filtered to remove frequency around $(2\omega_0 + \delta\omega)$, resulting in signals oscillating only at the offset frequency $\delta\omega$. This procedure is called *quadrature detection* and the outputs of such technique are two demodulated and filtered signals denoted $V_{\mathrm{P}}(t)$ and $V_{\mathrm{Q}}(t)$, oscillating at frequency around $\delta\omega$, that are put together in complex form

$$S(t) = V_{\mathrm{P}}(t) + iV_{\mathrm{Q}}(t), \tag{2.23}$$

where V_{P} represents the *in-phase signal* (real part) and V_{Q} the *quadrature signal* (imaginary part). After quadrature detection, the resulting NMR signal can be expressed as

$$S(t) \propto \omega_0 \int |B_+^r(\mathbf{r})||M_+(\mathbf{r},0)| e^{-t/T_2(\mathbf{r})} e^{-i(\delta\omega t + \phi_r(\mathbf{r}) - \phi_0(\mathbf{r}))} d\mathbf{r}. \tag{2.24}$$

From Eqs. (2.13b) and (2.22), we have

$$M_+(\mathbf{r},0) = |M_+(\mathbf{r},0)| e^{i\phi_0(\mathbf{r})}, \tag{2.25a}$$

$$\bar{B}_+^r(\mathbf{r}) = |B_+^r(\mathbf{r})| e^{-i\phi_r(\mathbf{r})}, \tag{2.25b}$$

where $\bar{B}_+^r$ denotes the complex conjugate of B_+^r. Eq. (2.24) can be further simplified to

$$S(t) \propto \omega_0 \int \bar{B}_+^r(\mathbf{r}) M_+(\mathbf{r},0) e^{-t/T_2(\mathbf{r})} e^{-i\delta\omega t} d\mathbf{r}. \tag{2.26}$$

Eq. (2.26) is here a function of T_2 only but the use of different combinations of RF pulses and gradient fields can condition the signal to be a function of the proton density, T_1 relaxation, T_2 relaxation, T_2^* relaxation and several other tissue properties.

2.4 Spatial localisation

Spatial localisation refers to applying additional spatial and time-varying magnetic fields on top of the static magnetic field, to allow spins to be excited at different frequencies at different locations, and hence making possible the representation of an object that is spatially inhomogeneous. These magnetic fields are referred to as *gradient magnetic fields* or simply as gradients. The use of gradients to spatially encode information was first proposed by Lauterbur in 1973 [4].

The first step of spatial localisation is to select the slice by using a gradient in the z-direction, referred to as *slice selective* gradient and denoted G_z. Once the RF pulse has been made spatially slice selective, the rest of the process is in-plane localisation and known as *spatial encoding*. This is achieved with a *phase-encoding* gradient in the y-direction (G_y) and a *frequency-encoding* (or *read-out*) gradient in the x-direction (G_x).

By defining the offset frequency $\delta\omega = \gamma(G_x x + G_y y + G_z z)$ (with γ the gyromagnetic ratio) in Eq. (2.26), it yields

$$S(t) \propto \omega_0 \int \bar{B}_+^r(\mathbf{r}) M_+(\mathbf{r},0) e^{-t/T_2(\mathbf{r})} e^{-i\gamma(G_x x + G_y y + G_z z)t} d\mathbf{r}. \tag{2.27}$$

Furthermore, if we introduce the vector notation

$$\mathbf{k}(t) = \frac{\gamma}{2\pi}\begin{bmatrix} G_x t \\ G_y t \\ G_z t \end{bmatrix} = \frac{\gamma}{2\pi}\big(G_x t\hat{\mathbf{i}} + G_y t\hat{\mathbf{j}} + G_z t\hat{\mathbf{k}}\big), \tag{2.28}$$

we can write the final expression as

$$S(\mathbf{k}) \propto \omega_0 \int \bar{B}_+^r(\mathbf{r}) M_+(\mathbf{r}, 0) e^{-t/T_2(\mathbf{r})} e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})} d\mathbf{r}. \tag{2.29}$$

In general, note that gradients also vary in time and should take the form

$$\mathbf{k}(t) = \frac{\gamma}{2\pi}\begin{bmatrix} \int_0^t G_x(t')\,dt' \\ \int_0^t G_y(t')\,dt' \\ \int_0^t G_z(t')\,dt' \end{bmatrix}. \tag{2.30}$$

The information about how gradient magnetic fields should behave in order to produce the trajectory $\mathbf{k}(t)$ is contained in the *pulse sequence*. This sequence also gives the number of RF pulse to generate the NMR signal: the basic types of pulse sequences in MRI are gradient-echo which uses a single RF pulse and spin-echo which uses two RF pulses.

2.5 Fourier encoding

Eq. (2.29) can be rewritten as

$$S(\mathbf{k}) = \int_{\mathbb{R}^D} I(\mathbf{r}) e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})} d\mathbf{r} = \mathcal{F}\{I(\mathbf{r})\}, \tag{2.31}$$

with the following assumptions: (i) the decay constant $T_2(\mathbf{r})$ is space-independent, (ii) the coil receives a spatially homogeneous magnetic field such that $\mathbf{B}^r = 1$, (iii) the scaling constant $\omega_0 e^{-t/T_2}$ is omitted, and (iv) the change of variable from $M_+(\mathbf{r}, 0)$ to $I(\mathbf{r})$. In Eq. (2.31), $I(\mathbf{r}) : \mathbb{R}^D \to \mathbb{C}$ represents the image function, $S(\mathbf{k}) : \mathbb{R}^D \to \mathbb{C}$ the MRI signal and $\mathcal{F}$ the Fourier integral transform operator. This expression makes explicit the relation between the image function (transverse magnetisation) and measured signal through the Fourier transform. This representation is commonly referred to as *Fourier encoding*, which stipulates that the detected NMR signals constitute a spatial frequency and phase representation of the object being imaged. This dual representation between the MR image and Fourier domain was described independently in the early 1980s by Ljunggren [5] and Twieg [6]. Note the frequency domain $S(\mathbf{k})$ is usually referred to as $\mathbf{k}$-space in MRI.

More generally, Eq. (2.31) can be expressed as

$$S(\cdot) = \int_{\mathbb{R}^D} I(\mathbf{r}) E(\mathbf{r}, \cdot) d\mathbf{r}, \tag{2.32}$$

where S, I and E refer respectively to the MRI signal, image and encoding functions. The encoding basis function takes the form $E(\mathbf{r}, \mathbf{k}) = e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})}$ in the case of Fourier encoding. Non-Fourier encoding methods have also been developed based on encoding basis functions such as wavelet [7,8] or singular value decomposition [9,10], but we will not discuss them in this dissertation due to the widespread of Fourier imaging in current MR technology.

Finally, it is important to emphasise at this point that both the NMR raw signal S and image function I are complex-valued. Although magnitude images are usually only displayed because it contains most of the relevant information, quantitative information can be obtained from the phase. For example, the phase can be used in cardiac imaging to assess cardiovascular flow measurement or in MR angiography to image velocity of moving blood.

2.6 Image reconstruction

The ideal goal of the MR image reconstruction problem would be to find the unknown *continuous* function I from the *discrete* measurement vector $\mathbf{y} \in \mathbb{C}^M$,

$$y_m = S(\mathbf{k}_m) = \int_{\mathbb{R}^D} I(\mathbf{r}) e^{-i2\pi(\mathbf{r}\cdot\mathbf{k}_m)} d\mathbf{r}, \quad m = 1, \ldots, M. \tag{2.33}$$

Of course, any *finite* set of Fourier samples cannot uniquely determine I because there are infinitely many feasible image functions that agree exactly with the given measured data [2,11].

There are multiple approaches to tackle problem (2.33) depending whether a continuous or discrete model is considered for both the data and object. A complete discussion is however out of scope of this dissertation; we refer the reader to the work of Fessler [11] for a thoughtful discussion. The continuous-data/continuous-object model is adopted hereinafter, which is the most common approach to explain the MR reconstruction problem. In this model, the hypothetical case of infinite sampling is considered (the function S is assumed to be known for all $\mathbf{k}$) to derive the analytical inversion formula, which is nothing else than the continuous inverse Fourier transform. Then, the fact the measurements can only be finite and discrete is taken into account, which leads to the inverse discrete Fourier transform.

To simplify the discussion in this section, only the one-dimensional case will be treated and the separability of the multidimensional Fourier transform is invoked to extend the procedure to multiple dimensions. Note that the discussion about

quantisation is also skipped. Quantisation is the post-sampling procedure that converts the measured values of the continuous function into a preassigned, finite set of number known to the computer. The reader is referred to Refs. [2, Chapter 6] and [12, Chapter 8] for more details.

2.6.1 Sampling

For the historical aspects, pioneering works on the sampling theory can be attributed independently to Kotelnikov, the Whittaker, Nyquist and Shannon in the first half of the 20^{th} century. In particular, the sampling theorem was introduced to communication engineers by Shannon in two seminal papers [13, 14] that founded the field of information theory. Today, it is most often known as the Nyquist-Shannon theorem. Its origins have been described in details in Ref. [15].

We assume uniform sampling in what follows. In one-dimension, Eq. (2.33) reads

$$S(k_n) = S(n\Delta k) = \int I(x) e^{-i2\pi(n\Delta k)x} dx, \tag{2.34}$$

where x represents the coordinate of spatial position, k the coordinate of spatial frequency and Δk is the sampling interval. The function $I(x)$ can be reconstructed from $S(n\Delta k)$ according to the Poisson summation formula,

$$\sum_{n=-\infty}^{\infty} S(n\Delta k) e^{i2\pi(n\Delta k)x} = \frac{1}{\Delta k} \sum_{n=-\infty}^{\infty} I\Big(x - \frac{n}{\Delta k}\Big). \tag{2.35}$$

This equation simply shows that the periodic summation of the function I (with period $1/\Delta k$) is completely defined by the Fourier coefficients $S(n\Delta k)$.

We now consider that the function $I(x)$ has bounded support, i.e. there exists a finite W such that $I(x) = 0$ for $|x| \geq W/2$. The region $|x| < W/2$ is known as the *field of view* (FOV) in MRI. To avoid aliasing artefacts, we assume that the Nyquist-Shannon sampling theorem is satisfied: if $I(x)$ has bounded support, its Fourier transform can be perfectly reconstructed from its sample values $S(n\Delta k)$ if

$$W < \frac{1}{\Delta k} \quad \text{or} \quad \Delta k < \frac{1}{W}. \tag{2.36}$$

It is said that the samples are taken at the *Nyquist rate* when $W = 1/\Delta k$ or $\Delta k = 1/W$. Sampling at a lower rate is called *undersampling*, while sampling at a higher rate is called *oversampling*. Since the Nyquist criterion is satisfied, there is no overlap in the various periodic replicas. Hence, the reconstruction formula for infinite sampling $\{n\Delta k, -\infty < n < +\infty\}$ can be written as

$$I(x) = \Delta k \sum_{n=-\infty}^{\infty} S(n\Delta k) e^{i2\pi(n\Delta k)x}, \quad |x| < \frac{1}{2\Delta k}. \tag{2.37}$$

where $I(x)$ is evaluated only within the FOV.

For finite sampling $\{n\Delta k, -N/2 \leq n < N/2\}$, the feasible reconstruction is not unique anymore. Assuming additional minimum norm constraint, the *Fourier reconstruction formula* for finite sampling can be derived which is in a form of a truncated Fourier series,

$$\widehat{I}(x) = \Delta k \sum_{n=-N/2}^{N/2-1} S(n\Delta k) e^{i2\pi(n\Delta k)x}, \quad |x| < \frac{1}{2\Delta k}. \tag{2.38}$$

$\widehat{I}$ is a (continuous) approximation of the true function I subject to ringing artefacts resulting from the truncation. Due to finite sampling, $I(x)$ is actually a band limited function, i.e. $S(k) = 0$ for $|k| > (N/2)\Delta k$. Therefore, recovering $I(x)$ from $\widehat{I}(m\Delta x)$ is possible only if the Nyquist criterion is satisfied,

$$\Delta x \leq \frac{1}{N\Delta k}. \tag{2.39}$$

If we choose $\Delta x = 1/(N\Delta k)$, we obtain

$$\begin{aligned} \widehat{I}(m\Delta x) &= \Delta k \sum_{n=-N/2}^{N/2-1} S(n\Delta k) e^{i2\pi(n\Delta k)(m\Delta x)} \\ &= \Delta k \sum_{n=-N/2}^{N/2-1} S(n\Delta k) e^{i2\pi nm/N}. \end{aligned} \tag{2.40}$$

With some modifications including the change of notations $\widehat{I}(m\Delta x) \to x_n$, $S(n\Delta k) \to y_k$, a shift in the index set and normalising with the factor $\Delta x = \Delta k = 1/\sqrt{N}$ to ensure that the transform is unitary, we obtain the one-dimensional inverse discrete Fourier transform (IDFT),

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} y_k e^{i2\pi kn/N}, \tag{2.41}$$

and the forward operation, the discrete Fourier transform (DFT),

$$y_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}. \tag{2.42}$$

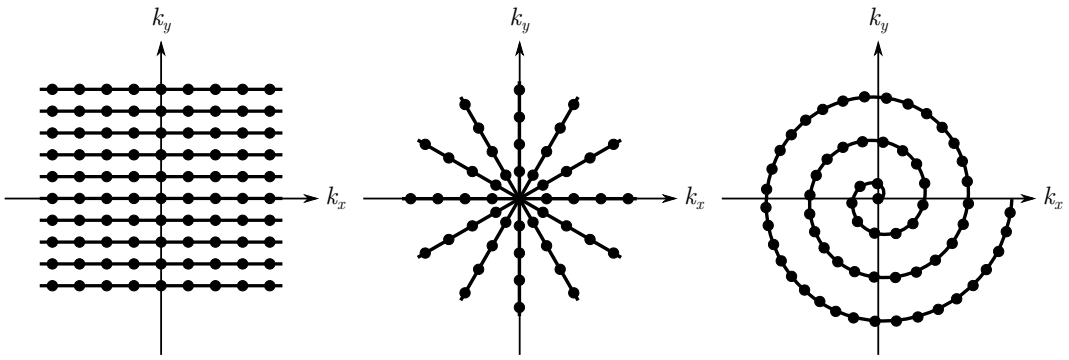

Figure 2.1: From left to right, standard Cartesian (rectilinear), radial and single shot spiral **k**-space trajectories in 2D.

2.6.2 Algebraic formulation

Eqs. (2.41) and (2.42) can be respectively expressed more compactly using a matrix-vector product

$$\mathbf{x} = \mathbf{F}^{-1}\mathbf{y}, \tag{2.43a}$$

$$\mathbf{y} = \mathbf{F}\mathbf{x}, \tag{2.43b}$$

where $\mathbf{F}$ represents the DFT matrix,

$$\mathbf{F} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{-i2\pi/N} & e^{-i4\pi/N} & \cdots & e^{-i2\pi(N-1)/N} \\ 1 & e^{-i4\pi/N} & e^{-i8\pi/N} & \cdots & e^{-i2\pi 2(N-1)/N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-i2\pi(N-1)/N} & e^{-i2\pi 2(N-1)/N} & \cdots & e^{-i2\pi(N-1)(N-1)/N} \end{bmatrix} \in \mathbb{C}^{N\times N}. \tag{2.44}$$

Since the matrix is unitary, we have $\mathbf{F}^{\mathrm{H}}\mathbf{F} = \mathbf{F}\mathbf{F}^{\mathrm{H}} = \mathbf{I}$ (where $\mathbf{I}$ is the identity matrix), and equivalently $\mathbf{F}^{-1} = \mathbf{F}^{\mathrm{H}}$. The advantages of formulation (2.43) is that it provides a convenient compact mathematical notation and the image reconstruction problem becomes one of solving a system of linear equations.

In practice the DFT and IDFT are actually never or rarely computed by explicitly defining such matrices as in Eq. (2.44). In fact, the N-points DFT and IDFT are most of the time not even computed using the naive definitions (2.42)-(2.41) that requires $\mathcal{O}(N^2)$ operations, but using the well known fast Fourier transform (FFT) algorithm that necessitates only $\mathcal{O}(N \log N)$ operations.

2.6.3 Cartesian and non-Cartesian sampling

The conventional way to acquire Fourier samples in MRI is along a uniformly sampled Cartesian space, also known as rectilinear sampling. It is also possible to collect

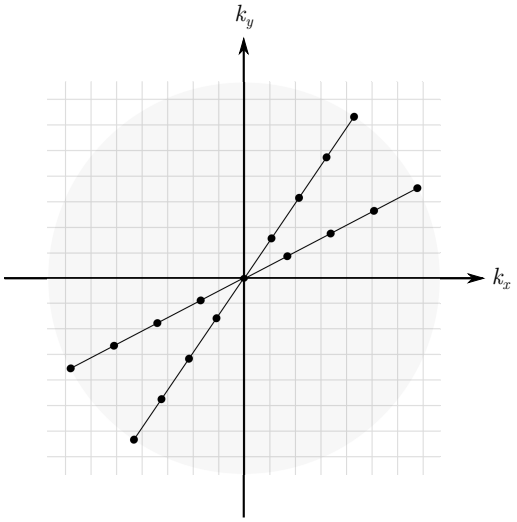

Figure 2.2: Non-Cartesian sampling illustrated with two radial projections. The problem is to generate uniform samples onto the Cartesian grid from nonuniformly spaced samples (black dots).

data on a non-Cartesian space by using radial or spiral trajectories as illustrated in figure 2.1. The **k**-space sampling trajectory is determined by the pulse sequence that contains information about gradient magnetic fields as discussed in section 2.4.

In non-Cartesian sampling, data are not collected on a rectangular grid which results in nonuniformly spaced samples acquired in the frequency domain. The problem of non-Cartesian reconstruction is then to generate uniformly space samples in the image domain. The gridding technique is a well established procedure that interpolates nonuniform samples onto a uniform rectilinear grid, generally via convolution of each data point with a Kaisser-Bessel convolution kernel. The gridding technique is a particular case of non-Cartesian reconstruction that treats the nonuniform (source) to the uniform (destination) case. Whether the source, destination or both data are nonuniform is a more general case that can be treated with the nonuniform Fast Fourier transform, see Refs. [16–19]. The non-Cartesian reconstruction problem is illustrated in figure 2.2.

Another possibility in case of radial sampling consists of considering image reconstruction from Radon transform samples, also known as image reconstruction from projections. Image reconstruction algorithms from Radon transform samples include

- direct backprojection and filtered back projection algorithms which approximate implementations of the inverse Radon transform,

- direct Fourier reconstruction that needs to convert the projection data to Fourier data using the projection slice theorem[1], interpolate the Fourier data

[1] The projection slice theorem stipulates that the one-dimensional Fourier transform of a pro-

to obtain a rectangular grid, and then use a standard Fourier reconstruction,

- algebraic reconstruction techniques where the image reconstruction is formulated as a set of algebraic equations.

In this dissertation, non-Cartesian sampling refers to *pseudo* non-Cartesian sampling in the sense that non-Cartesian trajectories will be directly approximated onto a Cartesian grid. This allows to simulate and test potentially many different types of sampling strategies without requiring actual non-Cartesian sampling from MR scans. A standard, real world, non-Cartesian acquisition would comprise a density compensation function that corrects for oversampling of the **k**-space centre and a gridding procedure to interpolate non-Cartesian data onto a Cartesian grid.

2.7 Noise

The principal source of noise in MRI is the thermal noise, also known as Johnson–Nyquist noise [20, 21], which is generated by random thermal agitation of charge carriers inside the receiver coil. Those random fluctuations primarily come from the patient's body, but also from the receiver coil itself and electronics [1, Chapter 15] [22].

Due to the nature of the thermal noise, the noise samples in **k**-space can be reasonably modelled by an additive normal distribution on both real and imaginary parts with independent and identically distributed random variables. The probability density function of the normal distribution reads

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.45}$$

with mean μ and variance σ^2, and is written more compactly as $\mathcal{N}(\mu, \sigma^2)$. Hence, instead of Eq. (2.31), the signal equation becomes

$$S(\mathbf{k}) = \int_{\mathbb{R}^D} I(\mathbf{r}) e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})} d\mathbf{r} + N(\mathbf{k}), \tag{2.46}$$

where N represents the noise function. For discrete data, the additive noise vector will be denoted by $\mathbf{n}$ and will obey

$$\forall j,\ \Re\{n_j\}, \Im\{n_j\} \sim \mathcal{N}(\mu, \sigma^2). \tag{2.47}$$

Since the Fourier transform is a linear transformation, the noise in a (complex-valued) MR image reconstructed through Fourier transform is also normally distributed. In fact, if normally distributed noise is assumed in each real and imaginary

jection taken at angle θ is equal to the central radial slice at angle θ of the two-dimensional Fourier transform of the original object.

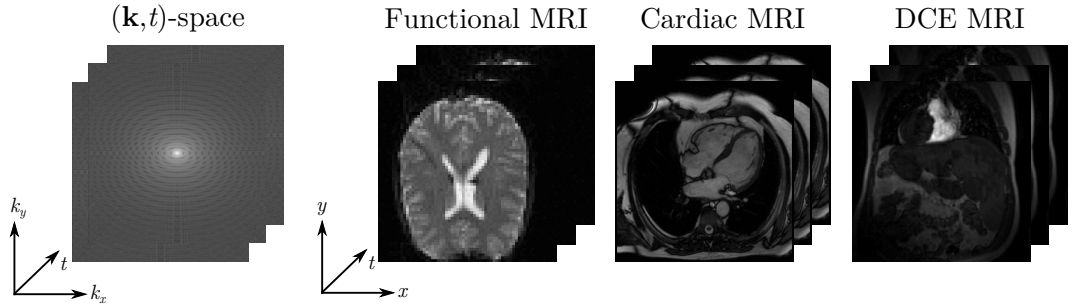

Figure 2.3: Dynamic magnetic resonance imaging. Left figure shows the **k**-space in time, often referred to as $(\mathbf{k}, t)$-space. Right figures represent common dynamic imaging data: functional, cardiac and dynamic contrast enhanced MRI.

parts, Gudbjartsson and Patz [23] showed that the noise in the magnitude of the MR image (absolute value of complex-valued data) were distributed according to the Rician distribution, whose probability density function is

$$f(x, \nu, \sigma) = \frac{x}{\sigma^2} \exp\left(\frac{-(x^2+\nu^2)}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right), \tag{2.48}$$

where x represents the measured pixel intensity, ν the image pixel intensity in the absence of noise, and $I_0(z)$ is the modified Bessel function of the first kind with order zero[2]. In regions where no NMR signal is present, they showed that the noise was governed by the Rayleigh distribution, a special case of the Rician distribution with $\nu = 0$. The general expression of the distribution for the phase image (argument of complex-valued data) is more complicated and omitted here.

Finally, note that Gudbjartsson and Patz [23] have also shown that a normal distribution of the noise for both magnitude and phase images is approximately valid when the *signal-to-noise ratio* (SNR) is larger than two. The SNR can be defined in the reconstructed image as the ratio of signal amplitude to the noise standard deviation, and it is a common measure to quantify the noise. In MRI, the SNR depends upon several imaging quality parameters such as spatial resolution and the number of acquired samples, and as such there is often a compromise to make between these parameters.

2.8 Dynamic imaging

The work presented in this dissertation focuses on dynamic MRI and the reconstruction of dynamic MR signals from sub-Nyquist sampling. Dynamic MR imaging refers to the acquisition of a series of MR images in time, resulting in a spatio-temporal MR signal where both spatial and temporal informations are available. Intuitively, the spatio-temporal signal can be seen as a sequence of images where dynamic events

[2] $I_0(z) = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+1)} \left(\frac{z}{2}\right)^{2m}$ where Γ is the Gamma function.

in the scene become visible much like a video, see figure 2.3. Dynamic MRI is an essential imaging modality to study and observe various dynamic phenomena.

Formally, dynamic MRI is based on an extension of the **k**-space definition with an additional time variable [24]. Instead of Eq. (2.46), the imaging equation can be written as

$$S(\mathbf{k}, t) = \int_{\mathbb{R}^D} I(\mathbf{r}, t) e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})} d\mathbf{r} + N(\mathbf{k}, t), \tag{2.49}$$

where S, I and N represent respectively the $(\mathbf{k}, t)$-space signal, the spatio-temporal image and noise functions.

2.9 Conclusion

We have given an overview of the core principles of the magnetic resonance imaging experiment in this chapter. The NMR phenomenon involves the alignment of the spin system in the presence of a static magnetic field, and the perturbation of this alignment by employing a second oscillating RF electro magnetic field at a specific frequency. The spin system relaxes back towards its original equilibrium state in the form of a rotating magnetisation that is detected and converted into an electrical signal via a receiver coil. Gradient magnetic fields are used to spatially encode information and make possible the formation of an image. Given a finite set of Fourier samples, image reconstruction is about finding a discrete approximation of the true continuous image function. Additional difficulties arise due to noise in MRI, that should be considered additive, complex-valued and normally distributed in **k**-space. Dynamic MR imaging involves a supplementary temporal dimension to image structures changing over time.

MRI is a sophisticated imaging modality, but the complete procedure is summarised in the technique's name. Indeed, *magnetic* refers to the interaction of nuclear magnetic moments in an assortment of magnetic fields; *resonance* relates to the matching of frequency between the RF pulse and the precession of the spins; finally, *imaging* refers to the process by which the signal is measured and then converted into an image.

Chapter 3

Linear inverse problems

Contents

3.1 Introduction

In this chapter, we discuss finite-dimensional linear inverse problems, a very common topic in many areas of science and engineering, and in particular in signal processing, imaging sciences and machine learning.

This chapter is organised as follows. We dicuss the general concepts of inverse problems in section 3.2. Signal recovery techniques from partial data are presented in section 3.3 as typical examples of inverse problems. We discuss state of the art methods for dynamic MRI reconstruction methods from undersampled (sub-Nyquist) data in section 3.4. A brief discussion on *inverse crimes* is provided in section 3.5 before concluding this chapter in section 3.6.

3.2 Generalities

3.2.1 Forward and inverse problems

Consider the finite-dimensional, linear, forward MRI model

$$\mathbf{y} = \mathbf{E}\mathbf{x}, \tag{3.1}$$

where $\mathbf{y} \in \mathbb{C}^M$ represents the $\mathbf{k}$-space measurements vector, $\mathbf{E} \in \mathbb{C}^{M\times N}$ is the MRI encoding matrix that models the acquisition procedure, and $\mathbf{x} \in \mathbb{C}^N$ is the signal/image to recover.

In this context, the *forward problem* is to find $\mathbf{y}$ given $\mathbf{x}$, that is to determine the observations from the object. The *inverse problem* is to find $\mathbf{x}$ given $\mathbf{y}$, or in other words to reconstruct the object from the measurements. There are three cases to consider to determine $\mathbf{x}$ from Eq. (3.1), depending whether $M = N$, $M > N$ or $M < N$.

1. $M = N$. The MRI encoding matrix models the sampling at Nyquist rate and Fourier operation, and is therefore the DFT matrix as in Eq. (2.44), i.e. $\mathbf{E} = \mathbf{F}$. The equation is invertible and the solution is given by $\mathbf{E}^{-1}\mathbf{y} = \mathbf{E}^{\mathrm{H}}\mathbf{y}$.

2. $M > N$. The MRI encoding matrix models over-Nyquist sampling and Fourier operation. In this case, there are more observations than unknowns. The system is said to be *overdetermined* and in general has no solution, although in some cases it may have a unique solution or infinitely many solutions. The *least squares* approach can be used to find an approximate solution (or exact solution if it exists) using the *Moore–Penrose pseudoinverse* $\mathbf{E}^{\dagger}$,

$$\begin{aligned} \mathbf{x}^{\star} &= \arg\min_{\mathbf{x}} \{F(\mathbf{x}) \equiv \|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2\} \\ &= \underbrace{(\mathbf{E}^{\mathrm{H}}\mathbf{E})^{-1}\mathbf{E}^{\mathrm{H}}}_{\mathbf{E}^{\dagger}}\mathbf{y}. \end{aligned} \tag{3.2}$$

   Indeed, we have $\nabla F(\mathbf{x}) = 2\mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{x} - \mathbf{y}) = \mathbf{0}$, and hence $\mathbf{x}^{\star} = (\mathbf{E}^{\mathrm{H}}\mathbf{E})^{-1}\mathbf{E}^{\mathrm{H}}\mathbf{y}$.

3. $M < N$. The MRI encoding matrix models sub-Nyquist sampling and Fourier operation. This can be expressed as $\mathbf{E} = \mathbf{A}\mathbf{F}$ where $\mathbf{A}$ represents the sampling

matrix (sub-Nyquist) and $\mathbf{F}$ the Fourier transform. In this case, there are less equations than unknowns. The system is said to be *underdetermined* and the solution of the system is not unique (infinitely many solutions). The *minimum norm* can be used when a system has multiple solutions,

$$\min_{\mathbf{x}}\{F(\mathbf{x}) \equiv \|\mathbf{x}\|_2^2\} \quad \text{s.t.} \quad \mathbf{y} = \mathbf{E}\mathbf{x}. \tag{3.3}$$

The solution of (3.3) can be derived via Lagrange multipliers and is given by

$$\mathbf{x}^\star = \underbrace{\mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{E}^{\mathrm{H}})^{-1}}_{\mathbf{E}^\dagger}\mathbf{y}. \tag{3.4}$$

Note from Eqs. (3.2) and (3.4) that the pseudoinverse $\mathbf{E}^\dagger$ has a dual definition.

3.2.2 Ill-posedness

We can characterise more formally the nature of an inverse problem with the following definition. A problem is said to be *well-posed* when it satisfies the following conditions [25–27]:

- a solution exists (*existence*),
- the solution is unique (*uniqueness*),
- the solution is stable with respect to small change in measurements (*stability*).

The problem is said to be *ill-posed* if it violates one or more of these requirements. The idea of a well-posed problem can be traced back to a short paper by Hadamard [28] in 1902 in the context of mathematical physics and boundary-value problems for partial differential equations.

From the above definition, it should be clear that the ideal MR image reconstruction problem (2.33), that is trying to find the true image function from a finite set of samples, is fundamentally an ill-posed inverse problem since there are infinitely many solutions. As for the finite-dimensional MRI reconstruction problem (3.1) from sub-Nyquist samples when $M < N$, this is a typical example of a linear discrete ill-posed inverse problem because there is no unique solution.

3.2.3 Optimisation and regularisation

Many inverse problems, and in particular MR image reconstruction problems, can be generally stated in the context of mathematical optimisation where an objective function $F(\mathbf{x}) : \mathbb{C}^N \to \mathbb{R}$ is minimised,

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}\in\mathbb{C}^N} F(\mathbf{x}). \tag{3.5}$$

When an inverse problem is ill-posed, a very common approach is to regularise the problem to reduce the space of candidate solutions. This translates into imposing additional constraints in the optimisation problem formulation. From a Bayesian point of view, this is equivalent to adding prior information based on a priori knowledge on the signal to reconstruct. This prior information can range from simple to more sophisticated knowledge of the signal one wish to recover. The objective function is then formulated as

$$\min_{\mathbf{x}}\{F(\mathbf{x}) \equiv \mathcal{D}(\mathbf{E}\mathbf{x}, \mathbf{y}) + \alpha\mathcal{R}(\mathbf{x})\}, \tag{3.6}$$

where $\mathcal{D}$ is a data fidelity criterion (observation model) and $\mathcal{R}$ represents the regularisation or penalty functional (prior information). The nonnegative parameter α is called *regularisation parameter* and controls the solution between the data fidelity term and the regularisation term. When $\alpha \to 0$ (resp. $\alpha \to \infty$), the solution favours the data fidelity term (resp. the regularisation term). The formulation (3.6) is often referred to as *Tikhonov regularisation* in the inverse problems literature. Equivalent constrained optimisation forms, under mild conditions, include *Morozov regularisation*

$$\min_{\mathbf{x}} \mathcal{R}(\mathbf{x}) \quad \text{s.t.} \quad \mathcal{D}(\mathbf{E}\mathbf{x}, \mathbf{y}) \leq \epsilon, \tag{3.7}$$

and *Ivanov regularisation*

$$\min_{\mathbf{x}} \mathcal{D}(\mathbf{E}\mathbf{x}, \mathbf{y}) \quad \text{s.t.} \quad \mathcal{R}(\mathbf{x}) \leq \gamma, \tag{3.8}$$

where ϵ and γ are nonnegative parameters.

In practice, optimisation problems are solved through iterative algorithms where the solution is updated iteratively starting from an initial guess because closed-form solutions do not exist in general. Even when a closed-form solution exists, it is often better to use an iterative method in terms of computational cost.

3.3 Signal recovery via low-dimensional models

In this thesis, we are mainly interested in the reconstruction of a discrete signal from only a limited amount of noisy Fourier samples. Without any additional information, this problem seems difficult, if not impossible.

However, the signal of interest often lies in a much lower dimensional space than its original domain. This is generally true because most often the signal presents some form of redundancy or inner structure that can be exploited, unless for example the signal is pure noise. Many techniques exist to reduce the dimension of signals, often referred to as *dimensionality reduction* methods, such as principal component analysis, factor analysis or sparse models. When the signal lies in a

lower-dimensional space, the recovery of the signal from only a small amount of data may become possible.

In the following, recovery methods that exploit signal low-dimensionality are described. We first introduce *compressed sensing* (CS), an efficient signal acquisition and reconstruction method for the recovery of sparse (or approximately sparse) signals. We then describe *low-rank matrix recovery*, *low-rank plus sparse decomposition* and *low-rank tensor recovery*.

3.3.1 Compressed sensing

Compressed sensing (CS) finds its origins in a paper by Candès et al. [29] published in 2006. They showed that it was possible to perfectly reconstruct a signal from a limited amount of Fourier samples, using a priori information about the sparsity of the signal. The same year, Donoho [30] showed that an exact reconstruction was possible not only in the Fourier domain but in any other transform that could sparsify the signal of interest. He introduced the term *compressed sensing* which literally means the acquisition (sensing) of compressed data. The fundamental concepts underlying the CS theory, namely sparsity, a specific sensing procedure and nonlinear reconstruction, are described hereinafter.

Sparsity

A N-dimensional signal $\mathbf{s}$ is said to be *sparse* if it has only a few nonzero K coefficients compared to its intrinsic dimension N, i.e. $K \ll N$. A signal is said *approximately sparse* when it has only a few significant nonzero K coefficients compared to its intrinsic dimension N. The latter definition is useful because most often the signal is not purely sparse. A simple and intuitive measure of the signal's sparsity is the ℓ_0 pseudonorm,

$$\|\mathbf{s}\|_0 = \#\{n : s_n \neq 0\}, \tag{3.9}$$

which counts the number of nonzero elements in $\mathbf{s}$ (see also appendix A.1).

A signal $\mathbf{x}$ might not be sparse naturally, but it may be possible to transform it into a domain where it can be considered sparse by using a *sparsifying transform*. Specifically, if we consider a unitary sparsifying transform $\mathbf{\Psi}$ that correctly sparsifies the signal $\mathbf{x}$, then we can write

$$\mathbf{s} = \mathbf{\Psi}\mathbf{x}, \quad \mathbf{x} = \mathbf{\Psi}^{\mathrm{H}}\mathbf{s}, \tag{3.10}$$

because the unitary property implies that

$$\mathbf{\Psi}^{\mathrm{H}}\mathbf{\Psi} = \mathbf{I}, \tag{3.11}$$

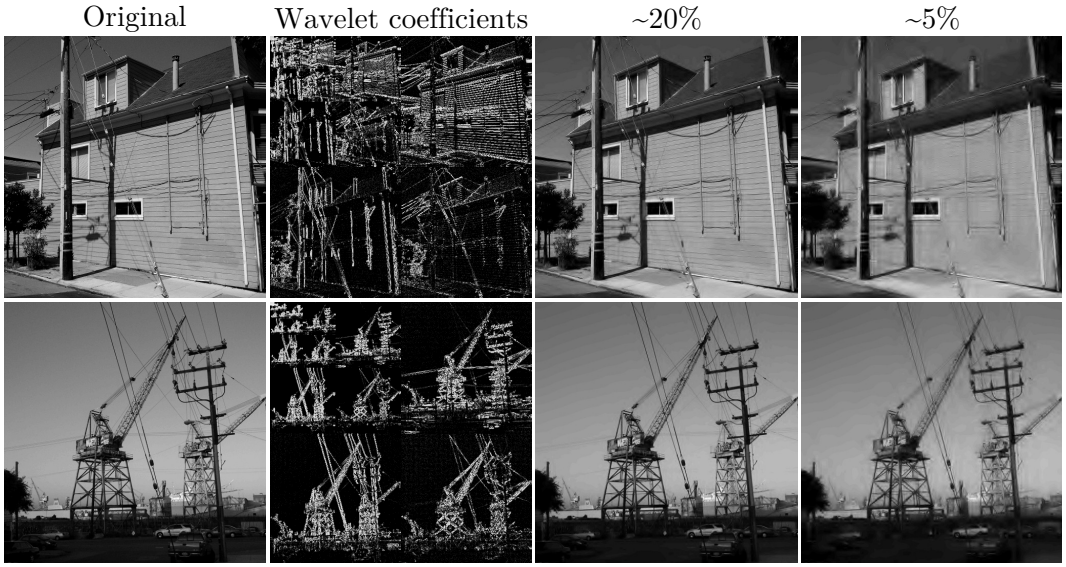

Figure 3.1: Two images converted into a sparse domain using a wavelet transform, and reconstructed with about 20% and 5% of their wavelet coefficients.

where $\mathbf{I}$ is the identity matrix. Sparsifying transforms are employed in image compression algorithms, such as JPEG2000 that uses a wavelet basis. Sparsity is related to compression because when a signal is sparse, only the nonzero coefficients and their locations need to be stored. Figure 3.1 illustrates the use of a wavelet basis as a sparsifying transform for images.

Sensing procedure

Consider the product of the sparse signal $\mathbf{s} \in \mathbb{C}^N$ with a sensing matrix $\mathbf{\Theta} \in \mathbb{C}^{M \times N}$ with $M \ll N$. The product results in the underdetermined system of linear equations

$$\mathbf{y} = \mathbf{\Theta}\mathbf{s}, \tag{3.12}$$

where $\mathbf{y}$ represents the M-dimensional measurement vector. Note that when we have $\mathbf{s} = \mathbf{\Psi}\mathbf{x}$, we should consider $\mathbf{\Theta}$ as the product between an $M \times N$ matrix $\mathbf{\Phi}$ and sparsifying transform matrix $\mathbf{\Psi}^{\mathrm{H}}$, i.e.

$$\mathbf{y} = \mathbf{\Theta}\mathbf{s} = \mathbf{\Phi}\mathbf{\Psi}^{\mathrm{H}}\mathbf{s} = \mathbf{\Phi}\mathbf{x}. \tag{3.13}$$

To ensure that compressed sensing algorithms perform well, the sensing matrix $\mathbf{\Theta}$ should either satisfy *incoherence* properties, *restricted isometry properties* (RIP) or *null space properties* (NSP). For example, the matrix $\mathbf{\Theta}$ should have a low coherence whose definition and bounds are

$$\mu(\mathbf{\Theta}) = \max_{1 \le i \ne j \le N} |\langle \boldsymbol{\theta}_i, \boldsymbol{\theta}_j \rangle|, \quad \mu(\mathbf{\Theta}) \in \left[\sqrt{\frac{N-M}{M(N-1)}}, 1 \right], \tag{3.14}$$

assuming $\boldsymbol{\Theta}$ has normalised columns. Alternatively, it is said that the matrix $\boldsymbol{\Theta}$ satisfy the RIP [31] of order K and isometry constant $0 \leq \delta_K^{\boldsymbol{\Theta}} < 1$ if

$$(1 - \delta_K^{\boldsymbol{\Theta}})\|\mathbf{s}\|_2^2 \leq \|\boldsymbol{\Theta}\mathbf{s}\|_2^2 \leq (1 + \delta_K^{\boldsymbol{\Theta}})\|\mathbf{s}\|_2^2 \tag{3.15}$$

for all K-sparse vectors $\mathbf{s} \in \mathbb{C}^N$. This will be futher discussed in section 3.3.3.

Nonlinear reconstruction

Since the system (3.12) is underdetermined, there are in general infinitely many solutions. The minimum norm can be used, but its solution is unlikely to be sparse. To constrain the solution to be sparse, the problem can be formulated by using the ℓ_0 pseudonorm,

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \boldsymbol{\Theta}\mathbf{s}. \tag{3.16}$$

However, the use of the ℓ_0 pseudonorm makes this minimisation problem computationally intractable as the vector dimension increases. In fact, it reduces to a combinatorial search with a complexity exponential in M, where the number of combination to test can be computed through the binomial coefficient.

To obtain a more tractable problem, a common technique known as *convex relaxation* [32] replaces the ℓ_0 pseudonorm with the ℓ_1 norm (see definitions in appendix A.1). The ℓ_1 norm is the closest convex function to the ℓ_0 pseudonorm. The modified constrained optimisation problem then becomes

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \boldsymbol{\Theta}\mathbf{s}. \tag{3.17}$$

When noise is considered in the model, the problem can be written with a mixture of ℓ_2 and ℓ_1 norms as

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}\|_2^2 \leq \epsilon, \tag{3.18}$$

where ε is a small number that is generally set to the noise level. The unconstrained (Lagrangian) form of (3.18) reads

$$\min_{\mathbf{s}} \frac{1}{2}\|\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{s}\|_1, \tag{3.19}$$

where α represents the regularisation parameter. Unconstrained minimisation problems use the Lagrangian to basically convert a constrained problem into an unconstrained one. Constrained and unconstrained problems are equivalent in the sense that for some specific values of ϵ and α, the two problems share the same solution although the relationship is unknown. Finally, we have so far only formulated the *synthesis* approach, where the aim is to reconstruct the sparse signal $\mathbf{s}$. We should

also mentioned the *analysis approach* which reconstructs $\mathbf{x}$ directly, although when the sparsifying transform is unitary as in Eq. (3.11), analysis and synthesis problems are similar.

Hence, various forms of CS optimisation problems exist depending on the constrained or unconstrained (Lagrangian) cases, and on the analysis or synthesis forms. They are summarised as follows,

$$\text{(Synthesis/Constrained)} \qquad \min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}\|_2^2 \leq \epsilon \tag{3.20a}$$

$$\text{(Synthesis/Unconstrained)} \qquad \min_{\mathbf{s}} \frac{1}{2}\|\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{s}\|_1 \tag{3.20b}$$

$$\text{(Analysis/Constrained)} \qquad \min_{\mathbf{x}} \|\boldsymbol{\Psi}\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{\Phi}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon \tag{3.20c}$$

$$\text{(Analysis/Unconstrained)} \qquad \min_{\mathbf{x}} \frac{1}{2}\|\boldsymbol{\Phi}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\boldsymbol{\Psi}\mathbf{x}\|_1. \tag{3.20d}$$

For the specific case of CS MRI [33,34], the sensing matrix corresponds to the MRI encoding matrix $\mathbf{E}$ that models both the random sub-Nyquist sampling and Fourier operation ($\mathbf{E} = \mathbf{AF}$), so that $\boldsymbol{\Theta} = \mathbf{E}\boldsymbol{\Psi}^{\mathrm{H}}$ and $\mathbf{y} = \mathbf{E}\boldsymbol{\Psi}^{\mathrm{H}}\mathbf{s} = \mathbf{E}\mathbf{x}$. We obtain these various forms of CS MRI optimisation problems,

$$\text{(Synthesis/Constrained)} \qquad \min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\mathbf{E}\boldsymbol{\Psi}^{\mathrm{H}}\mathbf{s} - \mathbf{y}\|_2^2 \leq \epsilon \tag{3.21a}$$

$$\text{(Synthesis/Unconstrained)} \qquad \min_{\mathbf{s}} \frac{1}{2}\|\mathbf{E}\boldsymbol{\Psi}^{\mathrm{H}}\mathbf{s} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{s}\|_1 \tag{3.21b}$$

$$\text{(Analysis/Constrained)} \qquad \min_{\mathbf{x}} \|\boldsymbol{\Psi}\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon \tag{3.21c}$$

$$\text{(Analysis/Unconstrained)} \qquad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\boldsymbol{\Psi}\mathbf{x}\|_1. \tag{3.21d}$$

The reconstruction procedure in CS involves a *nonlinear* method that enforces both sparsity and consistency with the acquired measurements, as opposed to a linear reconstruction by cardinal sine interpolation in standard Nyquist-Shannon sampling theory.

Finally, note that the use of ℓ_1 norm minimisation to obtain sparse signals can be traced back much earlier than in the mid-2000s for CS. It was used for geophysics problems in 1979 [35] and early 1980s [36,37], in the 1990s in the statistics community under the name *least absolute shrinkage and selection operator* (LASSO) [38], and in the signal processing community under the name *basis pursuit* [39].

3.3.2 Low-rank matrix recovery

An intuitive interpretation of the rank of a real or complex-valued matrix is given by the *singular value decomposition* (SVD) in the sense that a R-rank matrix will have exactly R nonzeros singular values. Any matrix can be decomposed using the SVD, which is a factorisation of a real or complex-valued $M \times N$ matrix $\mathbf{X}$ such

that

$$\mathbf{X} = \sum_{n=1}^{\min(M,N)} \sigma_n \mathbf{u}_n \mathbf{v}_n^{\mathrm{H}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}, \tag{3.22}$$

where $\mathbf{U}$ contains the left singular vectors (real or complex unitary square matrix), $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_n)$ is a diagonal matrix with nonnegative real-valued decreasing singular values σ_n on its diagonal, and $\mathbf{V}^{\mathrm{H}}$ contains the right singular vectors (real or complex unitary square matrix). The rank of the $M \times N$ matrix $\mathbf{X}$ can be considered low when

$$\mathrm{rank}(\mathbf{X}) \ll \min(M, N). \tag{3.23}$$

An important theorem that relates the rank and SVD is the Eckart-Young theorem [40] which states that the matrix

$$\mathbf{Y} = \sum_{n=1}^{K} \sigma_n \mathbf{u}_n \mathbf{v}_n^{\mathrm{H}}, \tag{3.24}$$

where $K \leq R$ is the optimal K-rank approximation of $\mathbf{X}$ in the Frobenius norm sense and assuming σ_n are decreasing singular values. Note that the SVD has a close relation with the Karhunen–Loève transform (KLT) and principal component analysis (PCA). In the context of 2D image processing, Gerbrands [41] showed that the three transforms are very similar if one single matrix is considered.

The low-rank matrix recovery problem is about reconstructing the low-rank matrix $\mathbf{X}$ from the measurement vector $\mathbf{y}$ according to

$$\mathbf{y} = \mathcal{A}(\mathbf{X}), \tag{3.25}$$

where the linear map $\mathcal{A} : \mathbb{C}^{M \times N} \to \mathbb{C}^P$ with $P \ll MN$ represents the sampling operator[1]. A naive approach is to consider the affine rank minimisation problem [42],

$$\min_{\mathbf{X}} \mathrm{rank}(\mathbf{X}) \quad \text{s.t.} \quad \mathbf{y} = \mathcal{A}(\mathbf{X}). \tag{3.26}$$

In words, one tries to find a matrix of minimum rank that satisfies a given system of linear equality constraints. However, the rank minimisation problem is computationally intractable as the dimension of the problem increases. Similarly to CS, a convex relaxation technique can be employed where the rank penalty is replaced by the nuclear norm, the closest convex surrogate of the matrix rank [42] which is defined as the sum of singular values (see appendix A.1),

$$\|\mathbf{X}\|_* = \sum_{n}^{\min(M,N)} \sigma_n. \tag{3.27}$$

[1]Note that the linear map $\mathcal{A}$ can be written with a matrix representation as $\mathcal{A}(\mathbf{X}) = \mathbf{A}\,\mathrm{vec}(\mathbf{X})$, where $\mathbf{A} \in \mathbb{C}^{P \times MN}$ and $\mathrm{vec}(\mathbf{X}) \in \mathbb{C}^{MN}$ denotes the vectorisation of $\mathbf{X}$.

3.3.3 Recovery guarantees

Recovery guarantees for sparse signal recovery can be obtained using the notion of coherence [43], restricted isometry properties [31], or null space properties [44]. Low-rank matrix recovery guarantees can be obtained using the generalisation of restricted isometry property to matrices [45] or null space conditions [46]. In what follows, we focus on recovery guarantees derived from the restricted isometry property.

Consider the model $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{x}$ a K-sparse vector and the sampling matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ $(M < N)$ and recall the RIP for sparse signal recovery. For integers $1 \leq K \leq N$, define the isometry constant $\delta_K^{\mathbf{A}}$ to be the smallest nonnegative number such that

$$(1 - \delta_K^{\mathbf{A}})\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K^{\mathbf{A}})\|\mathbf{x}\|_2^2 \tag{3.28}$$

for all K-sparse vectors $\mathbf{x}$.

The RIP was also generalised to matrices by Recht et al. [45]. Consider the model $\mathbf{y} = \mathcal{A}(\mathbf{X})$ where $\mathbf{X}$ is a R-rank matrix and $\mathcal{A} : \mathbb{R}^{M\times N} \to \mathbb{R}^P$ is a linear operator with $P < MN$. For integers $1 \leq R \leq \min(M, N)$, define the isometry constant $\delta_R^{\mathcal{A}}$ to be the smallest nonnegative number such that

$$(1 - \delta_R^{\mathcal{A}})\|\mathbf{X}\|_F^2 \leq \|\mathcal{A}(\mathbf{X})\|_F^2 \leq (1 + \delta_R^{\mathcal{A}})\|\mathbf{X}\|_F^2 \tag{3.29}$$

for all R-rank matrices $\mathbf{X}$.

Several conditions on the restricted isometry constants for exact recovery of both sparse signals and low-rank matrices have been proved by researchers. For example, the recovery of a K-sparse signal is guaranteed via ℓ_1 minimisation if the isometry constants of the measurement matrix $\mathbf{A}$ satisfy

$$\delta_{2K}^{\mathbf{A}} < \sqrt{2} - 1, \tag{3.30}$$

a sufficient condition proved by Candès [47] in 2008. More recently, Tai and Zhang [48] have established the following sharp conditions on the RIP. For sparse signal recovery, the theorem states that if the measurement matrix $\mathbf{A}$ satisfies

$$\delta_K^{\mathbf{A}} < \frac{1}{3} \tag{3.31}$$

for some integer $2 \leq K \leq N$, the minimiser of the constrained ℓ_1 norm minimisation problem,

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}}\{\|\mathbf{x}\|_1 : \mathbf{y} = \mathbf{A}\mathbf{x}\}, \tag{3.32}$$

recovers the K-sparse vector $\mathbf{x}$ exactly. For low-rank matrix recovery, the theorem

states that if the measurement operator $\mathcal{A}$ satisfies

$$\delta_R^{\mathcal{A}} < \frac{1}{3} \tag{3.33}$$

for some integer $2 \leq R \leq \min(M, N)$, the minimiser of the constrained nuclear norm minimisation problem,

$$\mathbf{X}^\star = \arg\min_{\mathbf{X}} \{\|\mathbf{X}\|_* : \mathbf{y} = \mathcal{A}(\mathbf{X})\}, \tag{3.34}$$

recovers the R-rank matrix $\mathbf{X}$ exactly. The above recovery guarantees hold for the noiseless case; the reader is referred to Ref. [48] for noisy signals.

However, coherence, RIP or NSP are not particularly convenient in practice. For example, the lower bound of the coherence in CS limits the performance analysis of recovery algorithms to relatively small sparsity levels as noted by Foucart and Rauhut [49]. Similarly, direct construction of a measurement matrix satisfying the RIP is computationally difficult due to the combinatorial nature of the problem. These limitations can be overcome by using random matrices. For example, Candès and Tao [50] have shown that the RIP in CS is satisfied with high probability for random Gaussian matrices of specific dimensions, i.e. if the entries of the matrix are independent Gaussian random variables. More generally, random matrices satisfy the RIP with high probability if their entries are chosen according to subgaussian distribution (which includes Gaussian and Bernoulli distributions). This has been shown independently by Mendelson et al. [51] and Baraniuk et al. [52]. For low-rank matrix recovery, it can also be shown that if linear mappings $\mathcal{A}$ are sampled from certain class of probability distributions (such as i.i.d. Gaussian), they satisfy the RIP for matrices with high probability [45].

3.3.4 Low-rank plus sparse matrix decomposition

The *low-rank plus sparse decomposition* (L+S) model [53–57], also sometimes referred to as *robust principal component analysis* (RPCA) or *principal component pursuit* (PCP) due to the paper by Candès et al. [56], is concerned with the exact decomposition of a given matrix into low-rank $\mathbf{L}$ and sparse $\mathbf{S}$ components. Considering a matrix $\mathbf{X}$, RPCA can be formulated as the following minimisation problem,

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{L} + \mathbf{S}. \tag{3.35}$$

This convex problem employs both the nuclear and ℓ_1 norms to respectively enforce a low-rank and a sparse matrix subject to data constraints.

A related problem referred to as *compressive principal component pursuit* is when

only partial observations are considered, i.e.

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathcal{A}(\mathbf{L}+\mathbf{S}), \tag{3.36}$$

where $\mathcal{A}$ represents a sampling operator. It has been shown in multiple studies [56, 58–62] that it is possible under some assumptions to recover both low-rank and sparse components from only a fraction of observations. The RPCA model will be further discussed in chapter 6.

3.3.5 Low-rank tensor recovery

A *tensor* $\boldsymbol{\mathcal{X}}$ is a multidimensional array of the form $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_Q}$ where Q represents the order of the tensor, also referred to as Q-mode or Q-way tensor. A second-order tensor is a matrix and a first-order tensor is a vector. A fiber of $\boldsymbol{\mathcal{X}}$ is a column vector defined by fixing every index of $\boldsymbol{\mathcal{X}}$ but one. Fibers are higher-order generalisation of matrix rows and columns.

The mode-q *unfolding* of a tensor $\boldsymbol{\mathcal{X}}$, also called *matricisation* or *flattening*, are denoted $\mathbf{X}_{(q)}$ and are obtained by arranging the mode-q fibers to be the columns of the resulting matrix. As an example, Kolda and Bader [63] consider a 3-mode tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ with $N_1 = 3$, $N_2 = 4$ and $N_3 = 2$. Consider that $\boldsymbol{\mathcal{X}}$ has the frontal slices

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 \end{bmatrix}. \tag{3.37}$$

The unfolding modes of $\boldsymbol{\mathcal{X}}$ are

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 13 & 14 & 15 & 16 \\ 5 & 6 & 7 & 8 & 17 & 18 & 19 & 20 \\ 9 & 10 & 11 & 12 & 21 & 22 & 23 & 24 \end{bmatrix}, \tag{3.38}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 5 & 9 & 13 & 17 & 21 \\ 2 & 6 & 10 & 14 & 18 & 22 \\ 3 & 7 & 11 & 15 & 19 & 23 \\ 4 & 8 & 12 & 16 & 20 & 24 \end{bmatrix}, \tag{3.39}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & \cdots & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & \cdots & 22 & 23 & 24 \end{bmatrix}. \tag{3.40}$$

However, the specific permutation of columns is generally not important as long as it is consistent across related calculations as Kolda and Bader [63] remark.

To define the rank of a tensor, we need first to introduce the definition of a *rank-*

one tensor. A Q-order tensor is rank-one if it can be written as the outer product of Q vectors,

$$\boldsymbol{\mathcal{X}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \ldots \circ \mathbf{a}^{(Q)}, \tag{3.41}$$

where $\circ$ is the vector outer product and $\mathbf{a}^{(q)} \in \mathbb{R}^{N_q}$, $q = 1, \ldots, Q$. The rank of tensor $\boldsymbol{\mathcal{X}}$ is defined as the smallest number of rank-one tensors that generate $\boldsymbol{\mathcal{X}}$ as their sum [63]. In mathematical terms

$$\operatorname{rank}(\boldsymbol{\mathcal{X}}) \equiv \min_{R \in \mathbb{N}} \Big\{ \exists \mathbf{a}_r^{(q)}, \tau_r \mid \boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \tau_r \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \ldots \circ \mathbf{a}_r^{(Q)} \Big\}, \tag{3.42}$$

where $\tau_r \in \mathbb{R}$ and $\mathbf{a}_r^{(q)} \in \mathbb{R}^{N_q}$ for $r = 1, \ldots, R$. For a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, Kolda and Bader [63] note that only the following weak upper bound on its maximum rank is known,

$$\operatorname{rank}(\boldsymbol{\mathcal{X}}) \leq \min\{N_1 N_2, N_1 N_3, N_2 N_3\}. \tag{3.43}$$

Kolda and Bader [63] also note that the rank of a real-valued tensor may be different over $\mathbb{R}$ and $\mathbb{C}$ and that there is no straightforward algorithm to determine the rank of a tensor.

For this reason, if is often considered instead the q-rank of tensors, which is also known as Tucker rank because it finds its foundation in the Tucker decomposition. The q-rank of $\boldsymbol{\mathcal{X}}$ is much easier to compute than $\operatorname{rank}(\boldsymbol{\mathcal{X}})$ because it is simply defined as the rank of the mode-q unfolding $\mathbf{X}_{(q)}$. In other words, the q-rank of Q-order tensor $\boldsymbol{\mathcal{X}}$ is the Q-dimensional vector whose q^{th} entry is the rank of the mode-q unfolding,

$$q\text{-rank}(\boldsymbol{\mathcal{X}}) \equiv \Big(\operatorname{rank}(\mathbf{X}_{(1)}), \ldots, \operatorname{rank}(\mathbf{X}_{(Q)}) \Big). \tag{3.44}$$

This definition implies that a low q-rank tensor is a tensor whose mode-q unfoldings are low-rank matrices. A more practical use of the q-rank is the sum of ranks of the mode-q unfoldings, i.e. $\sum_{q=1}^{Q} \operatorname{rank}(\mathbf{X}_{(q)})$.

Low-rank tensor recovery refers to a generalisation of low-rank matrix recovery to tensor, as higher-order generalisation of matrices. This topic has been mainly investigated through convex optimisation in the past [64–66]. The low-rank tensor recovery problem can be expressed as

$$\min_{\boldsymbol{\mathcal{X}}} \sum_{q=1}^{Q} \operatorname{rank}(\mathbf{X}_{(q)}) \quad \text{s.t.} \quad \mathbf{y} = \mathcal{A}(\boldsymbol{\mathcal{X}}), \tag{3.45}$$

where $\mathcal{A}$ is a sampling operator. Similar to sparse signal and low-rank matrix recovery, convex relaxation can be used to overcome intractability. The convex q-rank,

denoted c-q-rank, can be defined using the nuclear norm,

$$\text{c-}q\text{-rank}(\boldsymbol{\mathcal{X}}) = \sum_{q=1}^{Q} \|\mathbf{X}_{(q)}\|_*. \tag{3.46}$$

A tractable low-rank tensor recovery problem can then be written as

$$\min_{\boldsymbol{\mathcal{X}}} \sum_{q=1}^{Q} \alpha_q \|\mathbf{X}_{(q)}\|_* \quad \text{s.t.} \quad \mathbf{y} = \mathcal{A}(\boldsymbol{\mathcal{X}}), \tag{3.47}$$

where the α_q's represent weights. Problem (3.47) can be interpreted as the nuclear norm minimisations of multiple reorderings of the tensor into matrices.

3.4 Sub-Nyquist dynamic MRI

The motivations for accelerating MRI scans and in particular dynamic MRI have been explained in section 1.1. Many approaches have been proposed to reduce acquisition time from different perspectives. For example, echo planar imaging [67] and fast low-angle shot imaging [68] use specific pulse sequences, while parallel MR imaging uses multiple receiver coils.

In this thesis, the main approach we have taken to speed up scan time is based on violating the Nyquist criterion by skipping measurements that would be normally acquired in a standard MRI procedure. When the Nyquist criterion is not satisfied, we have seen that the discrete inverse problem is typically ill-posed because it lacks a unique solution. From an imaging aspect, this translates into the introduction of undesirable artefacts in the reconstructed images when the most basic reconstruction technique, the minimum norm, is used. Thus, the aim is to develop methods that can remove these artefacts and provide physiologically representative and accurate images in agreement with the measurements. In what follows, we briefly review some of these reconstruction techniques from undersampled $(\mathbf{k}, t)$-space that have been proposed so far.

3.4.1 Temporal and spatio-temporal interpolation

One of the first technique proposed to reconstruct dynamic MRI data from partial $(\mathbf{k}, t)$-space samples is the *sliding window* method. The most basic version of sliding window is the zeroth-order hold technique [69] which simply estimates the missing $\mathbf{k}$-space samples at time frame t with the latest data point at time frame $t-1$ from the same $\mathbf{k}$-space location.

A more sophisticated approach was proposed in 1999 called *unaliasing by Fourier-encoding the overlaps using the temporal dimension* (UNFOLD) [70]. This technique

uses a lattice undersampling scheme that produces aliasing artefacts that can be easily removed with simple filtering in the temporal Fourier domain.

In 2003, Tsao et al. [71] proposed an approach called *broad-use linear acquisition speed-up technique* (**k**-t BLAST). This method uses a variable-density sampling scheme, which is a combination of acquired random **k**-space lines (called undersampled dataset) with sampling of central part of the **k**-space (called training dataset). These datasets are converted into the temporal Fourier domain, and respectively contain aliasing artefacts and a low-spatial resolution estimate. **k**-t BLAST uses the training dataset to guide the reconstruction and removes artefacts in the undersampled dataset.

3.4.2 Compressed sensing

In 2006, Lustig et al. [72] proposed **k**-t SPARSE, a compressed sensing method for dynamic MR imaging that uses a sparsifying transform adapted for dynamic MR imaging, random undersampling and ℓ_1 norm reconstruction. **k**-t SPARSE solves the convex optimisation problem

$$\min_{\mathbf{x}} \|\mathbf{\Psi x}\|_1 \quad \text{s.t.} \quad \|\mathbf{Ex} - \mathbf{y}\|_2 \leq \epsilon, \tag{3.48}$$

where $\mathbf{\Psi x}$ represents the sparse signal with $\mathbf{\Psi}$ being a sparsifying transform, $\mathbf{E}$ is the MRI encoding operator modelling both the random sub-Nyquist sampling process and Fourier transform and $\mathbf{y}$ is the stacked $(\mathbf{k}, t)$-space measurements vector. The sparsifying transform $\mathbf{\Psi}$ represents in **k**-t SPARSE a Fourier transform in the temporal direction and a wavelet transform in the spatial direction. The temporal Fourier transform is further discussed in section 3.4.5. Formulation (3.48) is of Morozov type as shown in (3.7) with the constraint representing the data fidelity term. A nonlinear conjugate gradient descent algorithm with backtracking line search is used to solve the unconstrained (Lagrangian) version of problem (3.48). In this method described in Ref. [33], the absolute values of the ℓ_1 norm are approximated by smooth (differentiable) functions.

Later, Jung et al. proposed **k**-t FOCUSS [73–75] which is based on the *focal underdetermined system solver* (FOCUSS) [76], a general estimation method to find localised energy solution from limited data that employs successive quadratic optimisation to obtain sparse solutions. More specifically, **k**-t FOCUSS first estimates a low-resolution version of the (y-f)-space signal and performs a FOCUSS reconstruction to recover it. The (y-f)-space signal is the dynamic MRI signal in the temporal Fourier domain (see section 3.4.5) which proves to be sparse. Hence, **k**-t FOCUSS addresses the CS dynamic MRI problem (ℓ_1 minimisation) by recovering directly the dynamic signal in the temporal Fourier domain using a specific quadratic

optimisation technique. Consider the equation

$$\mathbf{y} = \mathbf{E}\mathbf{x} \tag{3.49}$$

where $\mathbf{y}$ and $\mathbf{x}$ represent respectively the undersampled $(\mathbf{k}, t)$-space measurement vector and the sparse $(y\text{-}f)$-space image, and $\mathbf{E} = \mathbf{A}\mathbf{F}_y\mathbf{F}_t$ is the MRI encoding operator modelling the random sub-Nyquist sampling ($\mathbf{A}$), the Fourier transform along the y direction ($\mathbf{F}_y$) and the temporal Fourier transform ($\mathbf{F}_t$). The solution of Eq. (3.49) is not unique, and the minimum norm solution is unlikely to give a sparse reconstruction. Standard CS dynamic MRI methods solve

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{E}\mathbf{x}\|_2 \leq \epsilon. \tag{3.50}$$

Instead, consider the weighted minimum norm problem

$$\begin{aligned} &\text{find } \mathbf{x} = \mathbf{W}\mathbf{q}, \\ \text{where } \mathbf{q}: \quad &\min_{\mathbf{q}} \|\mathbf{q}\|_2 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{E}\mathbf{W}\mathbf{q}\|_2 \leq \epsilon, \end{aligned} \tag{3.51}$$

and where $\mathbf{W}$ is a weighting matrix. In its Lagrangian (unconstrained) form, the problem can also be written

$$\min_{\mathbf{q}} \|\mathbf{y} - \mathbf{E}\mathbf{W}\mathbf{q}\|_2^2 + \lambda\|\mathbf{q}\|_2 \tag{3.52}$$

which has the following closed form solution

$$\mathbf{q}^\star = \mathbf{W}^{\mathrm{H}}\mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{E}^{\mathrm{H}} + \lambda\mathbf{I})^{-1}\mathbf{y}. \tag{3.53}$$

Hence, the solution of problem (3.51) is

$$\mathbf{x}^\star = \mathbf{W}\mathbf{q}^\star = \mathbf{\Lambda}\mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{H}} + \lambda\mathbf{I})^{-1}\mathbf{y} \tag{3.54}$$

where $\mathbf{\Lambda} = \mathbf{W}\mathbf{W}^{\mathrm{H}}$. Now consider $\mathbf{x} = \widetilde{\mathbf{x}} + \mathbf{W}\mathbf{q}$ where $\widetilde{\mathbf{x}}$ represents a low-resolution estimate of the $(y\text{-}f)$-space, then $\mathbf{k}$-t FOCUSS solves a slightly modified version of problem (3.52),

$$\min_{\mathbf{q}} \|\mathbf{y} - \mathbf{E}\widetilde{\mathbf{x}} - \mathbf{E}\mathbf{W}\mathbf{q}\|_2^2 + \lambda\|\mathbf{q}\|_2 \tag{3.55}$$

which has the closed form solution

$$\mathbf{x}^\star = \widetilde{\mathbf{x}} + \mathbf{\Lambda}\mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{H}} + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{E}\widetilde{\mathbf{x}}). \tag{3.56}$$

Hence, at each iteration, $\mathbf{k}$-t FOCUSS essentially updates the weighting matrix $\mathbf{W}^k$

according to

$$\mathbf{W}^k = \begin{bmatrix} |x_1^{k-1}|^p & 0 & \cdots & 0 \\ 0 & |x_2^{k-1}|^p & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & |x_N^{k-1}|^p \end{bmatrix}, \quad 1/2 \leq p \leq 1, \tag{3.57}$$

where

$$\mathbf{x}^{k-1} = [x_1^{k-1}, x_2^{k-1}, \ldots, x_N^{k-1}]^\top \tag{3.58}$$

is the N-dimensional (y-f)-space vector estimate at $k-1$, and then computes $\mathbf{x}^k$ according to Eq. (3.56). For a weighting matrix power factor of $p = 1/2$ in $\mathbf{W}$, the authors of **k**-t FOCUSS showed that the FOCUSS solution is asymptotically equivalent to the ℓ_1 minimisation. One of the major advantage of **k**-t FOCUSS over **k**-t SPARSE is that it is computationally much more efficient since it uses quadratic (ℓ_2 norm) optimisation technique.

An improvement over **k**-t SPARSE called **k**-t group SPARSE was proposed by Usman et al. [77] in 2011, which exploits the fact that the sparse coefficients in the temporal Fourier domain typically form a group structure. An overview of some of these previously described methods can be found in the review paper by Tsao and Kozerke [69].

3.4.3 Low-rank

More recently, researchers have also looked at exploiting low-rank property of matrices in dynamic MRI, instead of simply vector sparsity as in CS. Consider a sequence of N_t images of dimensions $N_x \times N_y$. Approaches based on low-rank matrix recovery are usually based on the formulation of the *Casorati matrix*, a matrix of dimensions $N_x N_y \times N_t$ whose columns represent vectorised MR images of the dynamic sequence [78, 79]. Formally if we denote $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_t}$ the N_t vectorised MR images, the Casorati matrix takes the form

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_t}] \in \mathbb{C}^{N_x N_y \times N_t}. \tag{3.59}$$

This matrix is very likely to be approximately low-rank with only a few significant singular values because of the high correlation between images in time.

Methods have been proposed in 2010 by Haldar and Liang [79] and by Zhao et al. [80] to consider low-rank constraints for dynamic MRI reconstruction problems. Both methods solve the following problem

$$\min_{\mathbf{X}} \|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \text{rank}(\mathbf{X}) \leq C \tag{3.60}$$

where $\mathcal{E} : \mathbb{C}^{N_x N_y \times N_t} \to \mathbb{C}^M$ is the MRI encoding operator modelling both the random sub-Nyquist sampling and Fourier transform with $M \ll N_x N_y N_t$. Both methods are based on the *incremented-rank PowerFactorization* (IRPF) algorithm [81] which is an alternating least squares approach that makes use of a matrix factorisation of $\mathbf{X}$ to enforce C-rank structure implicitly.

Note that in problem (3.60), we have used the linear operator notation $\mathcal{E}$ instead of $\mathbf{E}$. In this thesis, we will in general use linear operators (e.g. $\mathcal{E}, \Psi$) instead of matrices (e.g. $\mathbf{E}, \mathbf{\Psi}$) to lighten the notations since we will handle both vector and matrix variables. In both cases linear operators can be written with a matrix representation as explained in the footnote of section 3.3.2.

3.4.4 Low-rank and sparsity

During this thesis, the combination of both low-rank and sparsity has also received interests by other researchers. More specifically, Lingala et al. [82] proposed in 2011 **k**-t SLR (*sparsity and low-rank structure*). The idea was to jointly use a nonconvex Schatten "norm" for the low-rank part and the spatio-temporal total variation (TV) for the sparsity part. The following minimisation problem is considered in **k**-t SLR,

$$\min_{\mathbf{X}} \|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha \|\mathbf{X}\|_p + \beta \, \mathrm{TV}(\mathbf{X}), \tag{3.61}$$

where $\|.\|_p$ is the nonconvex Schatten "p-norm" with $p = 0.1$ (see appendix A.1) and $\mathrm{TV}(\mathbf{X})$ is the spatio-temporal TV operator. In mathematical terms, $\mathrm{TV}(\mathbf{X})$ is the ℓ_1 norm of the gradient in directions x, y and t approximated by finite differences,

$$\mathrm{TV}(\mathbf{X}) = \left\| \sqrt{|\nabla_x(\mathbf{X})|^2 + |\nabla_y(\mathbf{X})|^2 + |\nabla_t(\mathbf{X})|^2} \right\|_1. \tag{3.62}$$

To solve (3.61), the authors in Refs. [82, 83] proposed to majorise the penalty terms by quadratic functions of $\mathbf{X}$. A simplified objective function is then defined using the majorisations, and the minimisation of this objective function is performed via a three-step alternating scheme. The three steps involve respectively a quadratic optimisation problem, the generalised singular value soft thresholding operator for nonconvex rank constraint [84, 85] and the multidimensional shrinkage operator for the spatio-temporal total variation prior. In addition, a continuation strategy is implemented to overcome the trade-off between computational complexity and accuracy.

In 2012, Zhao et al. [86] proposed a technique named *partial separability-sparse* (PS-Sparse) to also exploit both the low-rank and sparsity a priori information using respectively the partial separability model and a temporal Fourier transform. A notable feature of this work lies in a single formulation of both constraints using a sparsity constraint to regularise the PS model, although the method do require

both the order of the PS model (rank) and the sparsity regularisation parameter. In PS-Sparse, the Casorati matrix is expressed as a matrix factorisation $\mathbf{X} = \mathbf{UV}$ where $\mathbf{U}$ represents a basis for spatial subspace of $\mathbf{X}$, and $\mathbf{V}$ represents a basis for temporal subspace of $\mathbf{X}$. The basis $\mathbf{V}$ is estimated prior to image reconstruction and the basis for spatial subspace is estimated via the minimisation problem

$$\min_{\mathbf{U}} \|\mathcal{E}(\mathbf{UV}) - \mathbf{y}\|_2^2 + \alpha \|\mathbf{UV}_f\|_1 \tag{3.63}$$

where $\mathbf{V}_f$ represents the temporal Fourier transform of each row of $\mathbf{V}$ and $\|\mathbf{UV}_f\|_1$ is the ℓ_1 norm of $\mathbf{UV}_f$ viewed as a vector. An algorithm based on half-quadratic regularisation [87, 88] with continuation is used to solve the optimisation problem (3.63).

More recently, Otazo et al. [89] proposed a reconstruction method for dynamic MRI using a low-rank plus sparse approach. The optimisation problem is

$$\min_{\mathbf{L},\mathbf{S}} \{F(\mathbf{L},\mathbf{S}) \equiv \|\mathbf{L}\|_* + \lambda \|\mathcal{F}_t(\mathbf{S})\|_1\} \quad \text{s.t.} \quad \mathcal{E}(\mathbf{L}+\mathbf{S}) = \mathbf{y}, \tag{3.64}$$

where $\mathcal{F}_t$ is the temporal Fourier transform operator. This method is an adaptation for dynamic MRI of the model discussed earlier in section 3.3.4. It considers a decomposition of $\mathbf{X}$ into a linear combination of a low-rank $\mathbf{L}$ and sparse $\mathbf{S}$ components, which is different than considering jointly $\mathbf{X}$ as low-rank *and* sparse as in $\mathbf{k}$-t SLR and PS-Sparse. Authors proposed a proximal gradient method to solve the unconstrained version of problem (3.64) by considering the objective function as $F\left(\begin{bmatrix}\mathbf{L}\\ \mathbf{S}\end{bmatrix}\right)$ instead of $F(\mathbf{L},\mathbf{S})$. We will further discuss this model in chapter 6 and the connections with our work [90] that was published simultaneously.

3.4.5 On the temporal Fourier transform

In a continuous setting, the temporal Fourier transform or similarly the Fourier transform along the time direction, can be defined as

$$I(\mathbf{r}, f) = \int I(\mathbf{r}, t) e^{-i2\pi t f} dt = \mathcal{F}_t\{I(\mathbf{r}, t)\}, \tag{3.65}$$

where $I(\mathbf{r}, t)$ represents the spatio-temporal image function. This operator is denoted by the symbol $\mathcal{F}_t$ and its matrix version by $\mathbf{F}_t$. The temporal Fourier domain is also often referred to as (x, f)-space or (y, f)-space.

In this thesis, we will mainly consider the temporal Fourier transform as sparsifying transform for dynamic MR imaging. The frequent use of this sparsifying transform in dynamic MR reconstruction methods [70–73] can be explained for the following reasons. To obtain a sparse signal in dynamic MRI, the temporal Fourier

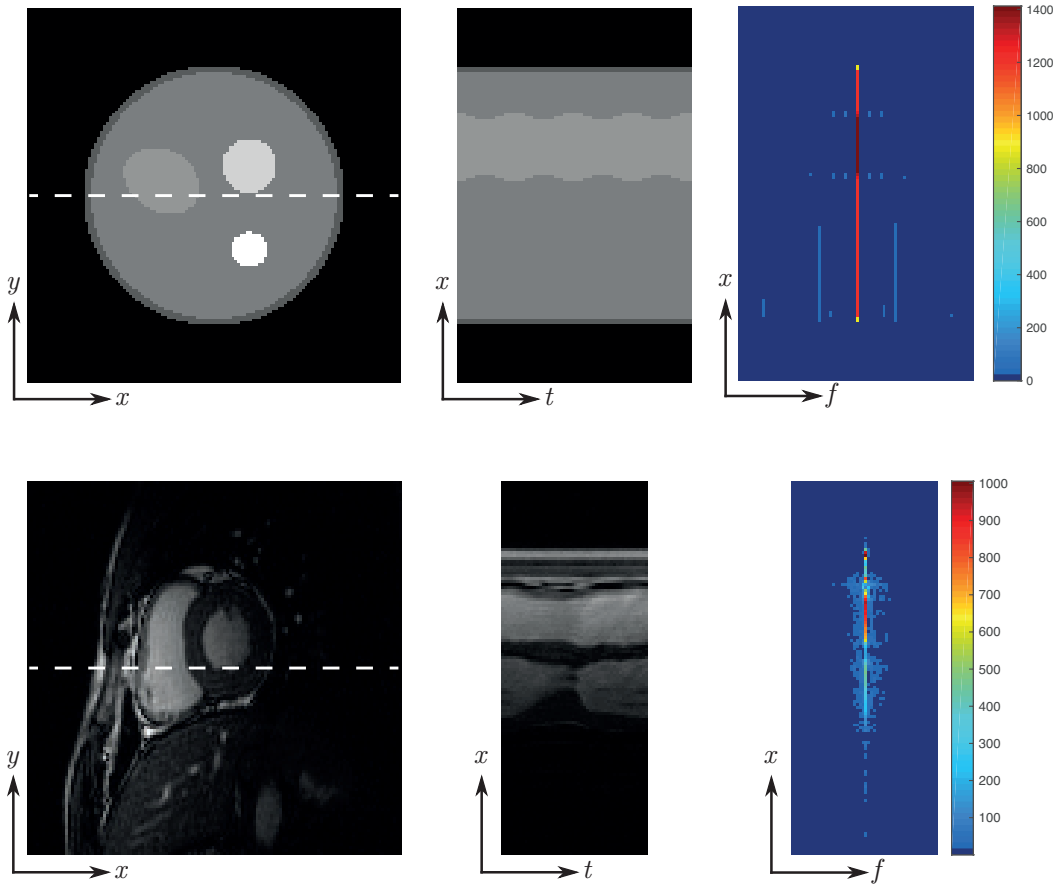

Figure 3.2: Illustration of the temporal Fourier transform for two typical dynamic MRI datasets. Time frame images from the sequence (left), temporal profiles along the dashed lines (middle) and Fourier transform along the temporal direction resulting in sparse (x-f)-space signals (right). Note the specific colour mappings to highlight the sparsity.

transform is particularly adequate because the spatio-temporal signal very often presents some periodicity in the time domain. In addition, this transform is unitary and it can be computed with the FFT algorithm which makes it one of the most efficient sparsifying transform in terms of computational cost. An illustration of a temporal Fourier transform is shown in figure 3.2.

3.5 Inverse crimes

Inverse crimes arise when the same model is used to generate the test data and to compute the reconstruction. They are intrinsically related to the process of converting an infinite-dimensional quantity to a finite-dimensional approximation necessitated by numerical processing. This has been highlighted in the inverse problem community [27, 91] and more recently in compressed sensing [92].

Inverse crimes are particularly widespread in the MR reconstruction literature

where the same discrete model is used both for simulation and reconstruction. Surprisingly, researchers in the MR community have only started recently to work out this problem. A notable approach has been proposed by Guerquin-Kern et al. [93] in 2012 who have developed an analytical phantom that can be defined with Bézier curve instead of rasterised image. This allows to take into consideration the continuous nature of data, providing more realistic reconstruction simulations.

Note that in this dissertation, we will also use the same discrete model for both simulation and reconstruction in numerical experiments. It should be acknowledged that it will result in artificially better reconstructions because the continuous nature of the data will not be taken into account.

3.6 Conclusion

In this chapter, we have reviewed linear discrete inverse problems. We have discussed important concepts such as ill-posedness, optimisation and regularisation in inverse problems. This chapter has mainly focused on signal recovery methods from partial data. These inverse problems can be solved using low-dimensional signal models by promoting low-complexity regularisation prior because most of the time signals lie in much lower dimensional spaces than their original domain. We have provided an overview of state of art reconstruction methods for sub-Nyquist dynamic MRI and briefly discussed inverse crimes.

Chapter 4

Optimisation framework: proximal splitting methods

Contents

4.1 Introduction

In this chapter, we introduce proximal splitting methods, a general framework to solve various convex optimisation problems.

Convex optimisation[1] [94–96] has increasingly gained in importance in recent years. One of the reason that explains this popularity is that even when the problem dimensions get large (roughly the number of variables and constraints), convex optimisation problems are still relatively easy to solve in contrast to nonconvex problems. In fact, although a common view is to generally interpret linear problems as easy and nonlinear ones as difficult, Rockafellar in his book *Lagrange multipliers and*

[1] We refer the reader to appendix A.2 for a brief review of some convex optimisation concepts.

optimality (1993) noted "In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and non-convexity". The popularity of convex optimisation methods can also be explained thanks to the availability of efficient minimisation algorithms to compute globally optimal solutions, the increase in computer power in the past decades (Moore's law), and the significant impact on the resolution of difficult problems when formulated with the help of convex relaxation (i.e. ℓ_1 and nuclear norm problems).

For these reasons, optimisation problems expressed as the minimisation of convex functionals are now ubiquitous in various areas such as in inverse problems [97], signal and image processing [98] or machine learning [99], and thus there is a huge interest in developing robust, fast and efficient algorithms for convex optimisation. However, difficulties arise when optimisation problems include nonsmooth and large-scale properties. Proximal splitting methods [100, 101] are first-order iterative algorithms for solving such convex optimisation problems. They operate by *splitting* the convex objective function to minimise which generates individual convex sub-problems. These sub-problems are evaluated easily via *proximal* operators, a generalisation of the projection operator. Proximal splitting methods offer a number of interesting properties that are particularly adapted for the work presented in this thesis:

- Convergence. Most of the proximal algorithms that we will describe in section 4.3 have convergence guarantees and/or potentially quantified competitive convergence rates.

- Computational speed. First-order methods are well suited for large-scale problems, mainly because iterations of typical first-order methods in the large-scale case remains cheap to evaluate (compared to interior-point methods for example). In the context of ℓ_1 and nuclear norms, first-order methods also possess nearly dimension-independent convergence rates as discussed by Nesterov and Nemirovski [102].

- Simplicity. These algorithms are in general short (a few lines) and easy to implement with minimal storage requirement.

- Flexibility. These algorithms can handle various general convex problems (potentially nonsmooth) and as such are flexible as long as the proximal operators can be evaluated easily. Due to the splitting approach, they also naturally fit the distributed and parallel computation framework [103].

This chapter is organised as follows. In section 4.2, we define the notion of proximal operator and give the analytical solutions of proximal operators relevant to this thesis. We review some proximal algorithms in section 4.3, and apply some of these algorithms in the context of compressed sensing MRI in section 4.4. We

also comment on nonconvex and greedy approaches in section 4.5, before concluding this chapter in section 4.6.

4.2 Proximal operators

4.2.1 Definition

The notion of proximal operator is best described by defining first the notion of the projection operator. The projection operator of $\mathbf{z}$ onto the closed convex subset $C \subseteq \mathbb{R}^N$, denoted $\text{proj}_C(\mathbf{z})$, is the solution to the following problem

$$\text{proj}_C(\mathbf{z}) = \arg\min_{\mathbf{x}\in C} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2. \tag{4.1}$$

Consider the characteristic function of C, defined as

$$\mathbf{1}_C(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad \mathbf{x} \in C \\ +\infty & \text{if} \quad \mathbf{x} \notin C \end{cases}, \tag{4.2}$$

which indicates the membership or nonmembership of a given element in that set. Then, the projection operator can be expressed with the characteristic function as

$$\text{proj}_C(\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + \mathbf{1}_C(\mathbf{x}). \tag{4.3}$$

The projection operator is used in the *projection onto convex sets* (POCS) algorithm to solve problems with simultaneous convex constraints. The POCS method is one of the first widely used convex optimisation technique and has been popularised in particular by Youla and Webb [104] in 1982.

The proximal (or proximity) operator was introduced by Moreau in 1962 [105] and further developed in 1965 [106]. He proposed to define the proximal operator of g, denoted $\text{prox}_g : \mathbb{R}^N \to \mathbb{R}^N$, as a generalisation of the convex projection operator. The characteristic function $\mathbf{1}_C$ in (4.3) is replaced by an arbitrary convex function g,

$$\text{prox}_g(\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + g(\mathbf{x}). \tag{4.4}$$

A proximal operator of g with parameter ρ can also be defined as

$$\begin{aligned} \text{prox}_{\rho g}(\mathbf{z}) &= \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + \rho g(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2\rho}\|\mathbf{x}-\mathbf{z}\|_2^2 + g(\mathbf{x}). \end{aligned} \tag{4.5}$$

Note that in this thesis, given the nature of complex-valued data in MRI, the proximal operator will be defined over convex functions of complex-valued variables, i.e.

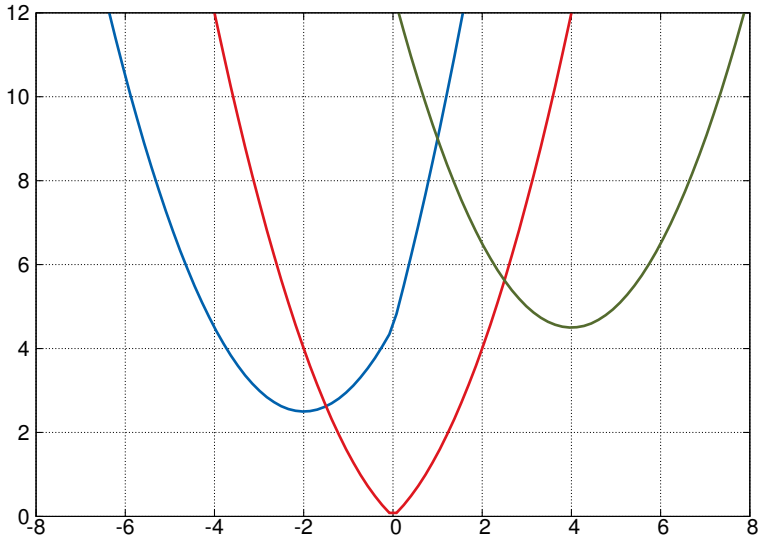

Figure 4.1: Graphs of functions $f(x) = \frac{1}{2}(x-z)^2 + \alpha|x|$ with $\alpha = 1$ and $z = -3$ (blue), $z = 0$ (red) and $z = 5$ (green). Using the soft thresholding operator as defined in Eq. (4.7), we obtain $\mathcal{S}_1^r(-3) = -2$, $\mathcal{S}_1^r(0) = 0$ and $\mathcal{S}_1^r(5) = 4$.

$\text{prox}_g : \mathbb{C}^N \to \mathbb{C}^N$.

4.2.2 Absolute value and ℓ_1 norm

Consider the proximal mapping of the absolute value function $g(x) = |x|$ with $x \in \mathbb{R}$,

$$\text{prox}_{\alpha|.|}(z) = \arg\min_{x\in\mathbb{R}} \left\{ \frac{1}{2}(x-z)^2 + \alpha|x| \right\}. \tag{4.6}$$

Although the absolute value is not differentiable at zero, a closed form solution for this proximal operator exists. It is given by the real-valued *soft thresholding* operator denoted $\mathcal{S}_\alpha^r$,

$$\mathcal{S}_\alpha^r(z) = \begin{cases} z-\alpha & \text{if} \quad z \geq \alpha \\ 0 & \text{if} \quad |z| \leq \alpha \\ z+\alpha & \text{if} \quad z \leq -\alpha \end{cases}. \tag{4.7}$$

This follows from the subdifferential of the absolute value function. This is illustrated in figure 4.1.

In case where $x \in \mathbb{C}$,

$$\text{prox}_{\alpha|.|}(z) = \arg\min_{x\in\mathbb{C}} \left\{ \frac{1}{2}|x-z|^2 + \alpha|x| \right\}, \tag{4.8}$$

the proximal operator is given by the complex-valued soft thresholding operator [49] denoted $\mathcal{S}_\alpha^c$,

$$\mathcal{S}_\alpha^c(z) = \begin{cases} \text{sgn}(z)(|z|-\alpha) & \text{if } |z| \geq \alpha \\ 0 & \text{otherwise} \end{cases}, \tag{4.9}$$

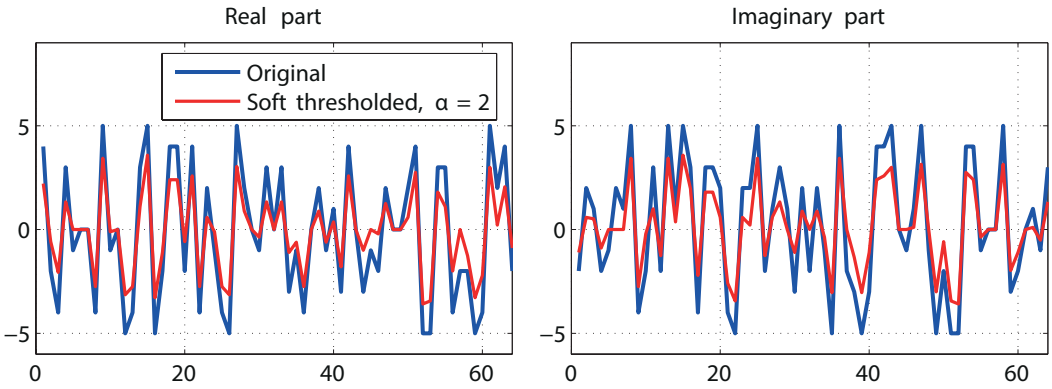

Figure 4.2: The soft thresholding operator applied on a complex-valued random signal. The resulting signal (in red) is shrunk, hence the alternative "shrinkage" name.

where sgn denotes the complex sign function,

$$\operatorname{sgn}(z) = \frac{z}{|z|}, \quad \forall z \in \mathbb{C}, \tag{4.10}$$

except for $z = 0 + i0$ where $\operatorname{sgn}(0 + i0) = 0$.

The soft thresholding operator can be defined more generally for the vector case with either real or complex-valued entries. We denote it simply $\mathcal{S}_\alpha$ and it is defined element-wise as

$$\begin{aligned} \mathcal{S}_\alpha(\mathbf{z}) = \{\mathcal{S}_\alpha(z_n)\}_{n=1}^N &= \frac{z_n}{|z_n|}\max(|z_n| - \alpha, 0) \\ &= \operatorname{sgn}(z_n)\max(|z_n| - \alpha, 0) \\ &= \operatorname{sgn}(z_n)(|z_n| - \alpha)_+, \end{aligned} \tag{4.11}$$

where sgn denotes the complex sign function. This operator thus defined is the solution to the proximal operator of the ℓ_1 norm,

$$\begin{aligned} \operatorname{prox}_{\alpha\|.\|_1}(\mathbf{z}) &= \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \alpha\|\mathbf{x}\|_1 \right\} \\ &= \mathcal{S}_\alpha(\mathbf{z}). \end{aligned} \tag{4.12}$$

Note the soft thresholding operator is also sometimes referred to as the *shrinkage operator*; see figure 4.2.

Finally, consider a unitary sparsifying transform operator Ψ (i.e. $\Psi^*\Psi = I$). Then, the proximity operator $\operatorname{prox}_{\alpha\|\Psi(.)\|_1}(\mathbf{z})$ can be defined as

$$\begin{aligned} \operatorname{prox}_{\alpha\|\Psi(.)\|_1}(\mathbf{z}) &= \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \alpha\|\Psi(\mathbf{x})\|_1 \\ &= \Psi^*\{\mathcal{S}_\alpha(\Psi(\mathbf{z}))\}. \end{aligned} \tag{4.13}$$

4.2.3 Nuclear norm

The proximal operator of the nuclear norm, the closest convex approximation of the rank function, has a closed-form solution given by the singular value (soft) thresholding (SVT) operator [107] that we denote $\text{SVT}^{\mathcal{S}}$,

$$\begin{aligned}\text{prox}_{\alpha\|.\|_*}(\mathbf{Z}) &= \arg\min_{\mathbf{X}\in\mathbb{C}^{M\times N}} \frac{1}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2 + \alpha\|\mathbf{X}\|_* \\ &= \text{SVT}^{\mathcal{S}_\alpha}(\mathbf{Z}) \\ &= \mathbf{U}\,\text{diag}(\mathcal{S}_\alpha(\text{diag}(\mathbf{\Sigma})))\mathbf{V}^{\text{H}},\end{aligned} \tag{4.14}$$

where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$ represents the singular value decomposition of $\mathbf{Z}$ and S_α is the soft thresholding operator defined as in Eq. (4.11). Note that in this case we have $\mathbf{X} \in \mathbb{C}^{M\times N}$, thus the proximal operator is defined over complex-valued matrix variables, i.e. $\text{prox}_{\alpha\|.\|_*} : \mathbb{C}^{M\times N} \to \mathbb{C}^{M\times N}$.

4.3 Proximal splitting algorithms

4.3.1 Proximal gradient methods

Sum of two convex functions

The following optimisation problem is first considered,

$$\min_{\mathbf{x}\in\mathbb{C}^N} F(\mathbf{x}) \equiv f(\mathbf{x}) + g(\mathbf{x}), \tag{4.15}$$

where it is assumed that $f : \mathbb{C}^N \to \mathbb{R}$ is a continuously differentiable (C^1) convex function and $g : \mathbb{C}^N \to \mathbb{R}$ is a convex function possibly nondifferentiable (see appendix A.3).

The *proximal gradient* method [108] to solve (4.15), also known as *forward-backward splitting*, is defined as the following iterative procedure

$$\mathbf{x}^{k+1} \leftarrow \text{prox}_{\rho g}(\mathbf{x}^k - \rho\nabla f(\mathbf{x}^k)), \tag{4.16}$$

where $\rho > 0$ is the step size (constant for all k or determined by line search). It is called forward-backward because at each iteration it uses the forward gradient step on f followed by a backward step on g, as the following decomposition suggests

$$\begin{aligned}\mathbf{x}^{k+1/2} &\leftarrow \mathbf{x}^k - \rho\nabla f(\mathbf{x}^k) \\ \mathbf{x}^{k+1} &\leftarrow \text{prox}_{\rho g}(\mathbf{x}^{k+1/2}).\end{aligned} \tag{4.17}$$

When ∇f is Lipschitz continuous with constant L (see appendix A.3), this method has been shown to converge for a fixed step size $\rho \in (0, 2/L)$, see Ref. [108]. Note that

the proximal gradient method can be seen as a generalisation of other algorithms: when $g = \mathbf{1}_C$, the method reduces to the projection onto C,

$$\mathbf{x}^{k+1} \leftarrow \mathrm{proj}_C(\mathbf{x}^k - \rho \nabla f(\mathbf{x}^k)), \tag{4.18}$$

and is known as *projected gradient method*; when $g = 0$, the method reduces to the standard *gradient descent*,

$$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^k - \rho \nabla f(\mathbf{x}^k). \tag{4.19}$$

Finally, it reduces to the *proximal point algorithm* when $f = 0$,

$$\mathbf{x}^{k+1} \leftarrow \mathrm{prox}_{\rho g}(\mathbf{x}^k), \tag{4.20}$$

which is also known as *proximal iteration*.

Since the proximal gradient method is slow in general, various methods have been proposed to accelerate it in particular by Nesterov [109] and Beck and Teboulle [110]. When the function f has a Lipschitz continuous gradient with constant L, these methods enjoy a fast rate of convergence on the objective function, i.e. $F(\mathbf{x}^k) - F(\mathbf{x}^\star)$ decreases at least as fast as $1/k^2$ with a fixed step size $\rho = 1/L$ or suitable line search, although the actual convergence of sequences produced by these schemes is no longer guaranteed [100]. The fast proximal gradient method of Beck and Teboulle [110] known as *fast iterative shrinkage-thresholding algorithm* (FISTA) is

$$\mathbf{x}^{k+1} \leftarrow \mathrm{prox}_{\rho g}(\mathbf{w}^k - \rho \nabla f(\mathbf{w}^k)) \tag{4.21a}$$

$$t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2}) \tag{4.21b}$$

$$\mathbf{w}^{k+1} \leftarrow \mathbf{x}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{x}^{k+1} - \mathbf{x}^k) \tag{4.21c}$$

where $\rho = 1/L$. In this case the major difference with the proximal gradient method is that the proximal step is not just used on the previous point $\mathbf{x}^k$, but at a point $\mathbf{w}^k$ that uses a specific linear combination of the previous two points $\{\mathbf{x}^{k+1}, \mathbf{x}^k\}$. Note that the specific steps (4.21b) and (4.21c) emerge from the analysis of the rate of convergence of FISTA. We refer the reader to Ref. [111] for more details. A simpler alternative identified by Vandenberghe [112] reads

$$\begin{aligned} \mathbf{x}^{k+1} &\leftarrow \mathrm{prox}_{\rho g}(\mathbf{w}^k - \rho \nabla f(\mathbf{w}^k)) \\ \mathbf{w}^{k+1} &\leftarrow \mathbf{x}^{k+1} + \frac{k-2}{k+1}(\mathbf{x}^{k+1} - \mathbf{x}^k) \end{aligned} \tag{4.22}$$

for $k \geq 1$.

Initially, FISTA was developed for ℓ_1 norm problems as a faster version of ISTA,

hence the presence of *shrinkage-thresholding* in the technique's name. The class of IST algorithms can be seen as an extension of the classical Donoho–Johnstone shrinkage method [113]. As described here and later on by the same authors in Ref. [111], the method can be generalised using the proximal formalism and is generally referred to as *fast proximal gradient* (FPG) or *accelerated proximal gradient* (APG) method.

Sum of multiple convex functions

We are now interested in the case of multiple convex functions with potentially more than two functions. The following optimisation problem is considered,

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}\in\mathbb{C}^N} \Big\{ F(\mathbf{x}) \equiv f(\mathbf{x}) + \sum_{j=1}^{J} g_j(\mathbf{x}) \Big\}, \tag{4.23}$$

where it is assumed that $f : \mathbb{C}^N \to \mathbb{R}$ is a continuously differentiable convex function and $g_j : \mathbb{C}^N \to \mathbb{R}$ are convex functions possibly nondifferentiables. The formulation of such objective functions enables the combination of multiple regularisation terms.

Huang et al. [114] have proposed in 2011 a method for minimising the sum of convex functions named *fast composite splitting* (FCS) algorithm. In practice, only a simplified version of FCS can be used which consists of the following iterations,

$$\mathbf{u}_j^{k+1} \leftarrow \mathrm{prox}_{J\rho g_j}(\mathbf{w}^k - \rho\nabla f(\mathbf{w}^k)) \tag{4.24a}$$

$$\mathbf{x}^{k+1} \leftarrow \frac{1}{J}\sum_{j=1}^{J} \mathbf{u}_j^{k+1} \tag{4.24b}$$

$$t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2}) \tag{4.24c}$$

$$\mathbf{w}^{k+1} \leftarrow \mathbf{x}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{x}^{k+1} - \mathbf{x}^k) \tag{4.24d}$$

assuming ∇f is L-Lipschitz continuous and $\rho = 1/L$. This algorithm averages proximal gradient steps in (4.24b) and contains an acceleration strategy borrowed from FISTA. While the above described algorithm is very similar to FISTA, it has not been strictly proved to have the same convergence rate as FISTA.

More recently, Raguet et al. [115] have introduced the *generalised forward-backward splitting* (GFBS) method that minimises the sum of multiple convex functions. The algorithm consists of the following iterations

$$\begin{aligned} \mathbf{w}_j^{k+1} &\leftarrow \mathbf{w}_j^k + \mathrm{prox}_{J\rho g_j}(2\mathbf{x}^k - \mathbf{w}_j^k - \rho\nabla f(\mathbf{x}^k) - \mathbf{x}^k) \\ \mathbf{x}^{k+1} &\leftarrow \frac{1}{J}\sum_{j=1}^{J} \mathbf{w}_j^{k+1}. \end{aligned} \tag{4.25}$$

When ∇f is L-Lipschitz, the method has been shown to converge when $\rho < 2/L$. As the name suggests it, this method can be seen as a generalisation of the forward-backward algorithm: for $J = 1$, this method simplifies to the forward-backward step (4.16).

4.3.2 Alternating direction method of multipliers

The *alternating direction method of multipliers* (ADMM) is a method for solving problems of the form

$$\min_{\mathbf{x},\mathbf{y}} \{F(\mathbf{x},\mathbf{y}) \equiv f(\mathbf{x}) + g(\mathbf{y})\} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{b}, \tag{4.26}$$

where both f and g are assumed to be convex functions and $\mathbf{A}$ and $\mathbf{B}$ are general matrices.

ADMM finds its origins in the mid-1970s from papers by Gabay and Mercier [116] and Glowinski and Marroco [117]. ADMM has regained interest in the past few years and has been used extensively in compressed sensing and inverse problems, see for example Refs. [118–122]. This method is in fact closely related to various other algorithms and has been rediscovered many times under different names. For example, split Bregman [118] has been shown to be very close to ADMM for linear constraints, and ADMM can be interpreted as the *Douglas-Rachford splitting* method on the dual. See the work of Esser [123] for further connections. ADMM can be seen in the more general framework of proximal splitting methods, and it proves to be a particularly suitable algorithm for matrix decomposition problems such as robust principal component analysis (RPCA) due to the fact that in RPCA a separable structure appears in both the objective function and the constraint. This will be further discussed in chapter 6. Finally, note that the convergence of ADMM has been proved in various papers. A summary of different proofs is given in Ref. [103].

The method is briefly described hereinafter, inspired by the review paper from Boyd et al. [103]. It is useful to recall the precursor techniques of *dual ascent* and the *method of multipliers* (MM) by considering first the linear equality-constrained optimisation problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{4.27}$$

where f is a convex function. Real-valued variables are considered to simplify notations.

Dual ascent

The associated Lagrangian function of problem (4.27) is

$$\mathcal{L}(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) + \mathbf{z}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}), \tag{4.28}$$

where $\mathbf{z}$ is the Lagrange multiplier. The dual problem consists of

$$\max_{\mathbf{z}}\{g(\mathbf{z}) \equiv \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{z})\}, \tag{4.29}$$

where $g(\mathbf{z})$ is the Lagrange dual function and inf denotes the infimum. The dual ascent method solves the dual problem using a gradient ascent step,

$$\mathbf{z}^{k+1} \leftarrow \mathbf{z}^k + \rho^k \nabla g(\mathbf{z}^k), \tag{4.30}$$

with ρ^k the step size and where $\nabla g(\mathbf{z}^k) = \mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}$ with

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{z}^k). \tag{4.31}$$

Hence, the dual ascent iterations reads

$$\mathbf{x}^{k+1} \leftarrow \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{z}^k) \tag{4.32a}$$

$$\mathbf{z}^{k+1} \leftarrow \mathbf{z}^k + \rho^k(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}). \tag{4.32b}$$

Method of multipliers

Consider now the *augmented Lagrangian* (AL) function of (4.27) which is the Lagrangian function augmented,

$$\mathcal{L}_\delta^A(\mathbf{x}, \mathbf{z}) = f(\mathbf{x}) + \mathbf{z}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\delta}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \tag{4.33}$$

where δ is called the penalty parameter. The *method of multipliers* (MM), also known as *augmented Lagrangian method* (ALM), consists in these iterations

$$\begin{aligned} \mathbf{x}^{k+1} &\leftarrow \arg\min_{\mathbf{x}} \mathcal{L}_\delta^A(\mathbf{x}, \mathbf{z}^k) \\ \mathbf{z}^{k+1} &\leftarrow \mathbf{z}^k + \delta(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}). \end{aligned} \tag{4.34}$$

Clearly, this is very close to the dual ascent algorithm, except that the augmented Lagrangian is used and the penalty parameter δ plays the role of the step size ρ^k. The MM can be interpreted as a more robust version of the dual ascent in terms of convergence.

ADMM

We now return to the original problem (4.26). The associated augmented Lagrangian function is

$$\mathcal{L}_\delta^A(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + g(\mathbf{y}) + \mathbf{z}^\top(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{b}) + \frac{\delta}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{b}\|_2^2. \tag{4.35}$$

The MM (or ALM) takes the following form,

$$(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) \leftarrow \arg\min_{\mathbf{x},\mathbf{y}} \mathcal{L}_\delta^A(\mathbf{x}, \mathbf{y}, \mathbf{z}^k) \tag{4.36a}$$

$$\mathbf{z}^{k+1} \leftarrow \mathbf{z}^k + \delta(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{b}), \tag{4.36b}$$

where in step (4.36a), the minimisation is performed simultaneously with respect to both $\mathbf{x}$ and $\mathbf{y}$. In contrast, ADMM consists in

$$\begin{aligned}
\mathbf{x}^{k+1} &\leftarrow \arg\min_{\mathbf{x}} \mathcal{L}_\delta^A(\mathbf{x}, \mathbf{y}^k, \mathbf{z}^k) \\
\mathbf{y}^{k+1} &\leftarrow \arg\min_{\mathbf{y}} \mathcal{L}_\delta^A(\mathbf{x}^{k+1}, \mathbf{y}, \mathbf{z}^k) \\
\mathbf{z}^{k+1} &\leftarrow \mathbf{z}^k + \delta(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{b}).
\end{aligned} \tag{4.37}$$

Clearly, the difference with the MM is that variables $\mathbf{x}$ and $\mathbf{y}$ are updated in an alternating way, hence the name *alternating direction method of multipliers*.

If we write (4.37) explicitly, we obtain

$$\begin{aligned}
\mathbf{x}^{k+1} &\leftarrow \arg\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + (\mathbf{z}^k)^\top \mathbf{A}\mathbf{x} + \frac{\delta}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}^k - \mathbf{b}\|_2^2 \right\} \\
\mathbf{y}^{k+1} &\leftarrow \arg\min_{\mathbf{y}} \left\{ g(\mathbf{y}) + (\mathbf{z}^k)^\top \mathbf{B}\mathbf{y} + \frac{\delta}{2}\|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y} - \mathbf{b}\|_2^2 \right\} \\
\mathbf{z}^{k+1} &\leftarrow \mathbf{z}^k + \delta(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{b}).
\end{aligned} \tag{4.38}$$

Usually, the linear and quadratic terms in the minimisations of the AL functions in (4.38) are combined to produce this more convenient form sometimes referred to as the *scaled form*,

$$\begin{aligned}
\mathbf{x}^{k+1} &\leftarrow \arg\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{\delta}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}^k - \mathbf{b} + \mathbf{u}^k\|_2^2 \right\} \\
\mathbf{y}^{k+1} &\leftarrow \arg\min_{\mathbf{y}} \left\{ g(\mathbf{y}) + \frac{\delta}{2}\|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y} - \mathbf{b} + \mathbf{u}^k\|_2^2 \right\} \\
\mathbf{u}^{k+1} &\leftarrow \mathbf{u}^k + \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{b},
\end{aligned} \tag{4.39}$$

where $\mathbf{u}^k = (1/\delta)\mathbf{z}^k$ is called the *scaled dual variable*.

Proximal version

In case where $\mathbf{A} = \mathbf{B} = \mathbf{I}$ in problem (4.26), we have

$$\begin{aligned}
\mathbf{x}^{k+1} &\leftarrow \arg\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{\delta}{2}\|\mathbf{x} - [\mathbf{b} - \mathbf{y}^k - \mathbf{u}^k]\|_2^2 \right\} \\
\mathbf{y}^{k+1} &\leftarrow \arg\min_{\mathbf{y}} \left\{ g(\mathbf{y}) + \frac{\delta}{2}\|\mathbf{y} - [\mathbf{b} - \mathbf{x}^{k+1} - \mathbf{u}^k]\|_2^2 \right\} \\
\mathbf{u}^{k+1} &\leftarrow \mathbf{u}^k + \mathbf{x}^{k+1} + \mathbf{y}^{k+1} - \mathbf{b}.
\end{aligned} \tag{4.40}$$

Using the definition (4.4) of the proximal operator, it is straightforward to derive the proximal version of the scaled form of ADMM,

$$\begin{aligned}\mathbf{x}^{k+1} &\leftarrow \text{prox}_{f/\delta}(\mathbf{b}-\mathbf{y}^k-\mathbf{u}^k)\\ \mathbf{y}^{k+1} &\leftarrow \text{prox}_{g/\delta}(\mathbf{b}-\mathbf{x}^{k+1}-\mathbf{u}^k)\\ \mathbf{u}^{k+1} &\leftarrow \mathbf{u}^k+\mathbf{x}^{k+1}+\mathbf{y}^{k+1}-\mathbf{b}.\end{aligned} \tag{4.41}$$

Interpretation as variable splitting plus AL method

ADMM is sometimes interpreted as a method based on variable splitting combined with the augmented Lagrangian method. Indeed, consider the minimisation of two convex functions

$$\min_{\mathbf{x}}\{F(\mathbf{x})\equiv f(\mathbf{x})+g(\mathbf{x})\}. \tag{4.42}$$

To solve this with ADMM, the problem is reformulated as a constrained one using the variable splitting procedure which simply consists in introducing a new variable,

$$\min_{\mathbf{x},\mathbf{y}}\{F(\mathbf{x},\mathbf{y})\equiv f(\mathbf{x})+g(\mathbf{y})\}\quad \text{s.t.}\quad \mathbf{x}=\mathbf{y}. \tag{4.43}$$

Based on the associated AL function of this modified problem,

$$\mathcal{L}_\delta^A(\mathbf{x},\mathbf{y},\mathbf{z})=f(\mathbf{x})+g(\mathbf{y})+\mathbf{z}^\top(\mathbf{x}-\mathbf{y})+\frac{\delta}{2}\|\mathbf{x}-\mathbf{y}\|_2^2, \tag{4.44}$$

the scaled form of ADMM reads

$$\begin{aligned}\mathbf{x}^{k+1} &\leftarrow \arg\min_{\mathbf{x}}\left\{f(\mathbf{x})+\frac{\delta}{2}\|\mathbf{x}-[\mathbf{y}^k-\mathbf{u}^k]\|_2^2\right\}\\ \mathbf{y}^{k+1} &\leftarrow \arg\min_{\mathbf{y}}\left\{g(\mathbf{y})+\frac{\delta}{2}\|\mathbf{y}-[\mathbf{x}^{k+1}+\mathbf{u}^k]\|_2^2\right\}\\ \mathbf{u}^{k+1} &\leftarrow \mathbf{u}^k+\mathbf{x}^{k+1}-\mathbf{y}^{k+1},\end{aligned} \tag{4.45}$$

where $\mathbf{u}^k=(1/\delta)\mathbf{z}^k$, and the proximal version is

$$\begin{aligned}\mathbf{x}^{k+1} &\leftarrow \text{prox}_{f/\delta}(\mathbf{y}^k-\mathbf{u}^k)\\ \mathbf{y}^{k+1} &\leftarrow \text{prox}_{g/\delta}(\mathbf{x}^{k+1}+\mathbf{u}^k)\\ \mathbf{u}^{k+1} &\leftarrow \mathbf{u}^k+\mathbf{x}^{k+1}-\mathbf{y}^{k+1}.\end{aligned} \tag{4.46}$$

4.4 Compressed sensing MRI example

In this section, we compare proximal gradient (PG), fast proximal gradient (FPG) and alternating direction method of multipliers (ADMM) in a simple compressed sensing MRI context.

Suppose we want to recover a sparse signal $\mathbf{s}\in\mathbb{C}^N$ from only partial Fourier sam-

ples. The complex-valued sparse signal $\mathbf{s}$ consists in randomly setting two nonzero values equal to one on both the real and imaginary parts. We construct the model $\mathbf{y} = \mathbf{E}\mathbf{s} + \mathbf{n}$ where $\mathbf{y} \in \mathbb{C}^M$ represents the observations corrupted by the complex-valued Gaussian noise vector $\mathbf{n}$ with zero mean and a standard deviation $\sigma = 0.01$. $\mathbf{E} \in \mathbb{C}^{M \times N}$ represents the random partial Fourier matrix and models the loss of information as well as the MRI encoding. We choose $N = 512$ and $M = 64$, i.e. the measurement vector $\mathbf{y}$ observes only 12.5% of the data. The optimisation problem consists in

$$\min_{\mathbf{s} \in \mathbb{C}^N} \Big\{ F(\mathbf{s}) \equiv \frac{1}{2} \|\mathbf{E}\mathbf{s} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{s}\|_1 \Big\}. \tag{4.47}$$

The proximal gradient to solve the sparsity-regularised problem (4.47) is the following simple step,

$$\mathbf{s}^{k+1} \leftarrow \mathrm{prox}_{\rho\alpha\|.\|_1}(\mathbf{s}^k - \rho \nabla f(\mathbf{s}^k)). \tag{4.48}$$

Algorithm 4.1 hence consists of only one line. Based on the proximal gradient algorithm, the fast proximal gradient can be easily derived by integrating the acceleration scheme as outlined in algorithm 4.2.

Algorithm 4.1 Compressed sensing MRI via proximal gradient

Input: $\mathbf{y}$, $\alpha \geq 0$, $\rho > 0$
Initialise: $k = 0$, $\mathbf{s}^0 = \mathbf{0}$
while stopping criterion is not met **do**
  $\mathbf{s}^{k+1} \leftarrow \mathcal{S}_{\rho\alpha}(\mathbf{s}^k - \rho \mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{s}^k - \mathbf{y}))$
end while
Output: $\widehat{\mathbf{s}} = \mathbf{s}^k$

Algorithm 4.2 Compressed sensing MRI via fast proximal gradient

Input: $\mathbf{y}$, $\alpha \geq 0$, $\rho > 0$
Initialise: $k = 0$, $\mathbf{s}^0 = \mathbf{w}^0 = \mathbf{0}, t^0 = 1$
while stopping criterion is not met **do**
  $\mathbf{s}^{k+1} \leftarrow \mathcal{S}_{\rho\alpha}(\mathbf{w}^k - \rho \mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{w}^k - \mathbf{y}))$
  $t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2})$
  $\mathbf{w}^{k+1} \leftarrow \mathbf{s}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{s}^{k+1} - \mathbf{s}^k)$
end while
Output: $\widehat{\mathbf{s}} = \mathbf{s}^k$

In ADMM, the problem must be reformulated by introducing a new variable,

$$\min_{\mathbf{s},\mathbf{b} \in \mathbb{C}^N} \Big\{ F(\mathbf{s}, \mathbf{b}) \equiv \frac{1}{2} \|\mathbf{E}\mathbf{s} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{b}\|_1 \Big\} \quad \text{s.t.} \quad \mathbf{s} = \mathbf{b}. \tag{4.49}$$

The associated AL function is

$$\mathcal{L}_\delta^A(\mathbf{s},\mathbf{b},\mathbf{z}) = \frac{1}{2}\|\mathbf{E}\mathbf{s}-\mathbf{y}\|_2^2 + \alpha\|\mathbf{b}\|_1 + \Re\{\mathbf{z}^{\mathrm{H}}(\mathbf{s}-\mathbf{b})\} + \frac{\delta}{2}\|\mathbf{s}-\mathbf{b}\|_2^2, \tag{4.50}$$

and the scaled form of ADMM reads

$$\mathbf{s}^{k+1} \leftarrow \arg\min_{\mathbf{s}} \left\{ \frac{1}{2}\|\mathbf{E}\mathbf{s}-\mathbf{y}\|_2^2 + \frac{\delta}{2}\|\mathbf{s}-\mathbf{b}^k+\mathbf{u}^k\|_2^2 \right\} \tag{4.51a}$$

$$\mathbf{b}^{k+1} \leftarrow \arg\min_{\mathbf{b}} \left\{ \frac{\alpha}{\delta}\|\mathbf{b}\|_1 + \frac{1}{2}\|\mathbf{s}^{k+1}-\mathbf{b}+\mathbf{u}^k\|_2^2 \right\} \tag{4.51b}$$

$$\mathbf{u}^{k+1} \leftarrow \mathbf{u}^k + \mathbf{s}^{k+1} - \mathbf{b}^{k+1}. \tag{4.51c}$$

By developing the analytical expressions in (4.51), we obtain algorithm 4.3.

Algorithm 4.3 Compressed sensing MRI via alternating direction method of multipliers

Input: $\mathbf{y}$, $\alpha \geq 0$, $\delta > 0$
Initialise: $k=0$, $\mathbf{s}^0 = \mathbf{b}^0 = \mathbf{u}^0 = \mathbf{0}$
while stopping criterion is not met **do**
  $\mathbf{s}^{k+1} \leftarrow (\mathbf{E}^{\mathrm{H}}\mathbf{E} + \delta\mathbf{I})^{-1}(\mathbf{E}^{\mathrm{H}}\mathbf{y} + \delta(\mathbf{b}^k - \mathbf{u}^k))$
  $\mathbf{b}^{k+1} \leftarrow \mathcal{S}_{\alpha/\delta}(\mathbf{s}^{k+1} + \mathbf{u}^k)$
  $\mathbf{u}^{k+1} \leftarrow \mathbf{u}^k + \mathbf{s}^{k+1} - \mathbf{b}^{k+1}$
end while
Output: $\widehat{\mathbf{s}} = \mathbf{s}^k$

The step size ρ must be chosen as $\rho \in (0, 2/L)$ for the proximal gradient, and $\rho = 1/L$ for the fast proximal gradient. As shown in appendix (A.3), the Lipschitz constant is $L = 1$, so that the step size is chosen as $\rho = 1$ for both PG and FPG. For ADMM, the penalty parameter is chosen as $\delta = 1$ which similarly plays the role of the step size. The regularisation parameter is set to $\alpha = 0.01$. Algorithms are stopped if a maximum number of 200 iterations is reached, or if the corresponding objective function F does not decrease significantly any more, i.e. when $[F(\mathbf{s}^{k+1}) - F(\mathbf{s}^k)]/F(\mathbf{s}^k) \leq 10^{-7}$.

Convergence of algorithms is shown in figure 4.3, and signal reconstructions are shown in figure 4.4. Table 4.1 provides some other quantitative results. The computational time for ADMM is significantly larger than PG and FPG due to the step (4.51a) which solves a system of linear equations explicitly in this implementation. Note that the behaviour of these algorithms changes depending on the regularisation parameter α, the step size ρ and the penalty parameter δ.

4.5 On nonconvex optimisation and greedy approaches

While convex optimisation has been adopted in this thesis for signal recovery, nonconvex approaches have been shown to provide superior solutions in some cases, e.g.

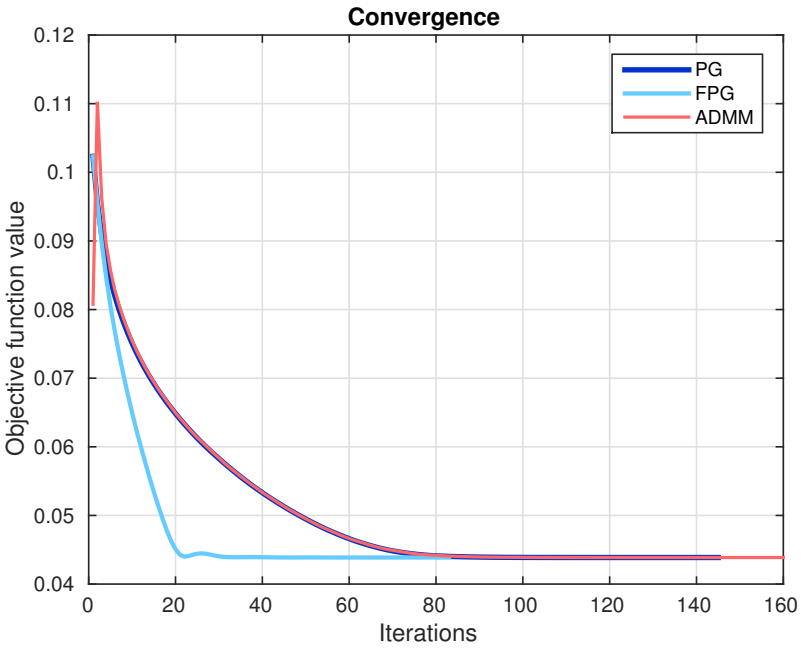

Figure 4.3: Convergence of algorithms.

	PG	FPG	ADMM
Iterations	145	83	160
Time (sec)	0.06	0.04	3.51
Abs. error	5.71×10^{-2}	5.69×10^{-2}	5.69×10^{-2}

Table 4.1: Number of iterations, execution times and absolute reconstruction errors for the CS algorithms.

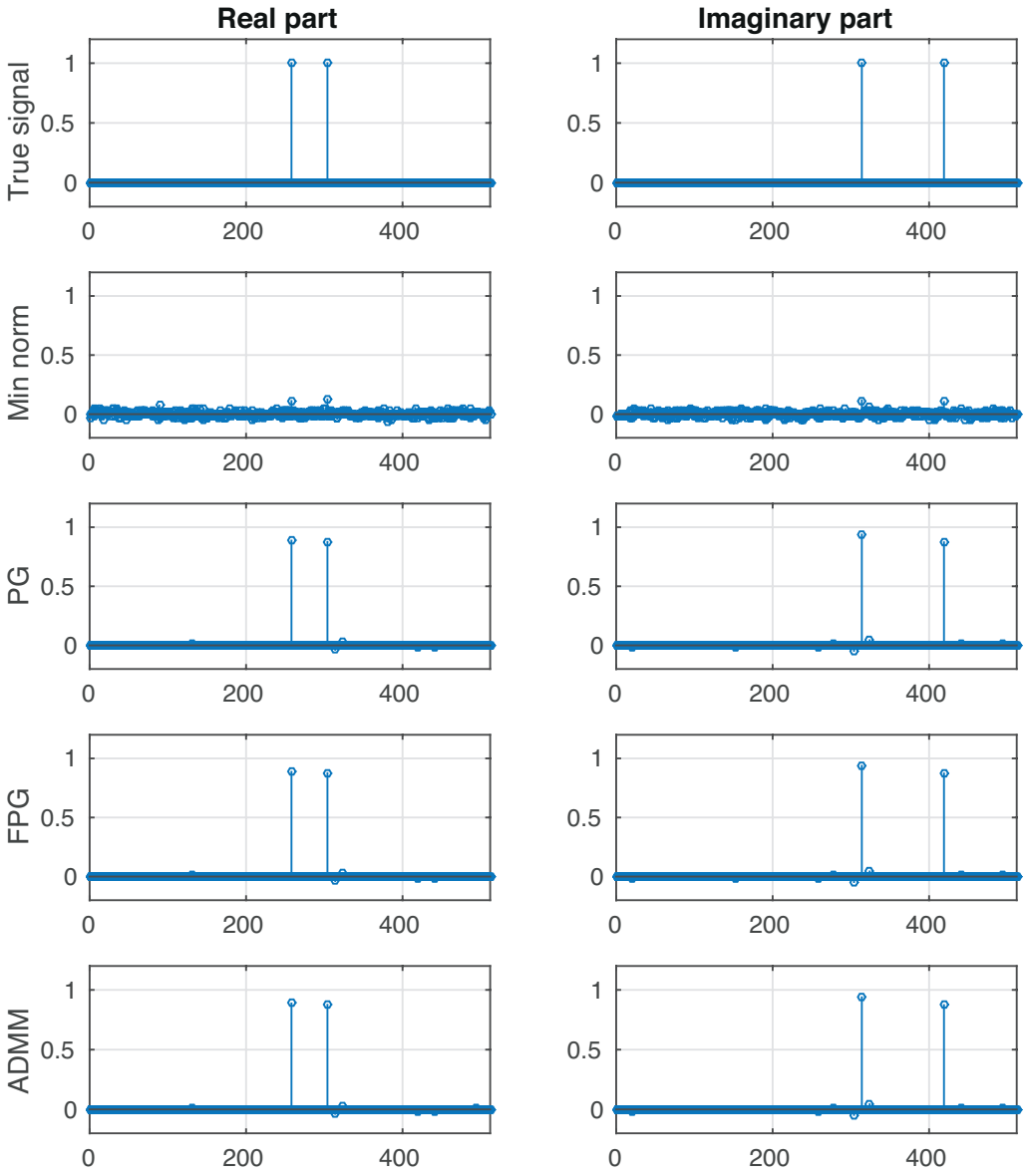

Figure 4.4: Original sparse signal, minimum norm solution as in Eq. (3.3) and the compressed sensing reconstructions (PG, FPG, ADMM). Both real and imaginary parts are shown as this is a complex-valued signal.

see Refs. [124] in CS, Refs. [84, 125] in CS MRI and Ref. [85] for the low-rank plus sparse model. For example in Ref. [124], Chartrand showed that fewer measurements were required to compute a sparse solution when using the nonconvex ℓ_p norm with $p < 1$ instead of $p = 1$. Trzasko et al. [125] have shown similar results in CS MRI based on homotopic approximation of the ℓ_0 norm. Major disadvantages of these methods are that achieving a global minimum is never guaranteed, and that they generally require higher computational costs than standard convex optimisation.

An alternative to convex or nonconvex optimisation principles is the greedy approach. Greedy methods operate directly on the signal coefficients in an iterative manner. For example, a basic greedy method is *orthogonal matching pursuit* [126]. Some greedy methods are particularly attractive in signal recovery from partial data because they also provide theoretical recovery guarantees that can rival with convex optimisation methods. Two important algorithms of this class include *iterative hard thresholding* (IHT) [127, 128] and *compressive sampling matching pursuit* (CoSaMP) [129]. IHT algorithms make use of the *hard thresholding* operator $\mathcal{H}_\alpha$, a nonlinear operator that sets all but the α largest entries (in magnitude) to zero,

$$\mathcal{H}_\alpha(\mathbf{z}) = \{\mathcal{H}_\alpha(z_n)\}_{n=1}^N = \begin{cases} z_n & \text{if } |z_n| > \alpha \\ 0 & \text{otherwise} \end{cases} . \tag{4.52}$$

For example, considering the low-rank matrix completion problem (3.26) and defining the singular value hard thresholding operator as

$$\mathrm{SVT}^{\mathcal{H}_\alpha}(\mathbf{Z}) = \mathbf{U}\mathcal{H}_\alpha(\mathrm{diag}(\mathbf{\Sigma}))\mathbf{V}^{\mathrm{H}}, \tag{4.53}$$

where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ represents the SVD of $\mathbf{Z}$, a simple IHT algorithm called *singular value projection* [130] is

$$\mathbf{X}^{k+1} \leftarrow \mathrm{SVT}^{\mathcal{H}_{\rho\alpha}}(\mathbf{X}^k - \rho\mathcal{A}^*(\mathcal{A}(\mathbf{X}^k) - \mathbf{y})), \tag{4.54}$$

where ρ is a constant step size. Other greedy approaches for low-rank matrix recovery with performance guarantees include *atomic decomposition for minimum rank approximation* (ADMiRA) [131] and normalised IHT [132]. For the low-rank plus sparse decomposition problem, Waters et al. [59] have proposed *sparse and low rank decomposition via compressive sensing* (SpaRCS) that combines the aspects of CoSaMP and ADMiRA.

4.6 Conclusion

In this chapter, we have outlined the optimisation framework used in this thesis, which is based on convex optimisation and the proximal splitting framework. We

have introduced the definition of the proximal operator and given the solution of proximal operators that are of interests in this thesis, namely low-rank and sparsity problems. We have described in details two important proximal algorithms, proximal gradient methods and alternating direction method of multipliers. We have provided a basic example of these algorithms in a compressed sensing setting. We finally have mentioned nonconvex and greedy alternatives in the last section.

Chapter 5

Fast proximal gradient methods

Contents

5.1 Introduction

This chapter develops and characterises computational tools based on proximal gradient methods to reconstruct dynamic MR images from a limited amount of noisy Fourier samples using sparse and rank deficiency assumptions.

The use of sparsity prior has been largely investigated in MRI in the last past decade due to the compressed sensing theory. More recently, exploiting low-rank structure has been proposed as an alternative to sparse approaches following important theoretical and algorithmic developments in low-rank matrix recovery. The combination of both low-rank and sparse assumptions has also attracted interest in the MR community [82, 83, 86, 89, 133, 134], where problems are formulated as the minimisation of an objective function that includes multiple penalty terms.

Classical optimisation techniques can be employed to minimise the corresponding objective functions, but it generally requires their modification to avoid nondifferentiability and other problems. Although leading to feasible schemes, these techniques often result in computationally-intensive optimisation problems. This chapter is motivated by a gap in the literature in providing simple and flexible algorithms to solve sparse, low-rank, and jointly sparse and low-rank problems in MRI. The choice of proximal splitting methods and convex optimisation was justified in chapter 4. The choice of proximal gradient methods over the alternating direction method of multipliers (ADMM) in this chapter is mainly motivated by the following facts,

- in the objective functions, the data fidelity term is differentiable and can be evaluated at little cost, so this information should be exploited,
- fast optimisation scheme with attractive rate of convergence have been developed for proximal gradient methods (i.e., FISTA [110], Nesterov's scheme [109] as described in section 4.3.1),
- as opposed to ADMM, there is no need to reformulate problems by introducing new variables and/or to establish the associated augmented Lagrangian function,
- although ADMM can also be developed for more than two convex functions, multiple penalty terms naturally fit proximal gradient methods.

The major contribution of this work is to develop and evaluate fast (convex) proximal gradient algorithms for sparse, low-rank and simultaneously sparse and low-rank priors for dynamic MRI reconstruction. In this work, the temporal Fourier transform is used as the sparsifying transform, and low-rank constraints are promoted via nuclear norm as described in more details in chapters 3 and 4. In addition, the developed algorithms are compared with three state of the art methods that were described in 3.4, **k**-t FOCUSS with temporal average [73–75] that exploits

sparsity in the temporal Fourier domain and two state of the art techniques that exploit both low-rank and sparsity, **k**-t SLR [82] and PS-Sparse [86]. In the process of developing these algorithms for low-rank and sparsity problems, we eventually provide a method that not only consistently offers higher reconstruction accuracy, but also proves to be highly competitive in terms of computational times. Another interesting aspect of this work is to assess reconstruction algorithms in the context of different types of datasets (phantom, invivo) and various characteristics such as complex-valued, noisy, motion-included and undersampled data, which are rarely evaluated all together in the literature. Parts of this chapter have been previously published in Refs. [135, 136].

This chapter is organised as follows. Section 5.2 describes the general methodology, that is proximal gradient algorithms and the sampling strategy employed to satisfy both MR and signal recovery constraints. Section 5.3 presents numerical simulations with a description of the various datasets used in this chapter and the reconstruction results. We discuss various aspects in section 5.4, related works in section 5.5 and conclude this chapter in section 5.6.

5.2 Method

5.2.1 Proximal gradient optimisation

We are interested in this chapter in minimisation problems of the form

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \Big\{ F(\mathbf{x}) \equiv f(\mathbf{x}) + \sum_{j=1}^{J} g_j(\mathbf{x}) \Big\}, \tag{5.1}$$

where the objective function F is formulated as composite convex functions with data fidelity term f and regularisation priors g_j. The function f is assumed to be a differentiable convex function with L-Lipschitz continuous gradient (see appendix A.3), and functions g_j are assumed to be convex but not necessary differentiable.

5.2.2 Sparse signal recovery

We consider a vector $\mathbf{x} \in \mathbb{C}^{N_x N_y N_t}$ that consists of N_t images of dimension $N_x \times N_y$. The finite-dimensional forward MRI model adopted is

$$\mathbf{y} = \mathcal{E}(\mathbf{x}) + \mathbf{n}, \tag{5.2}$$

where $\mathbf{y} \in \mathbb{C}^M$ is the stacked $(\mathbf{k}, t)$-space measurements vector, $\mathcal{E} : \mathbb{C}^{N_x N_y N_t} \to \mathbb{C}^M$ represents the MRI encoding operator modelling both the random sub-Nyquist sampling process and Fourier transform with $M \ll N_x N_y N_t$ and $\mathbf{n} \in \mathbb{C}^M$ is the complex-valued Gaussian noise vector.

We consider first the sparsity-regularised inverse problem

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}\in\mathbb{C}^{N_x N_y N_t}} \Big\{ F(\mathbf{x}) \equiv \frac{1}{2}\|\mathcal{E}(\mathbf{x}) - \mathbf{y}\|_2^2 + \alpha\|\mathcal{F}_t(\mathbf{x})\|_1 \Big\}, \tag{5.3}$$

where $\mathcal{F}_t$ represents the temporal Fourier transform as described in section 3.4.5. This can be solved efficiently with a simple proximal gradient step,

$$\mathbf{x}^{k+1} \leftarrow \text{prox}_{\rho\alpha\|\mathcal{F}_t(.)\|_1}(\mathbf{x}^k - \rho\nabla f(\mathbf{x}^k)). \tag{5.4}$$

Developing this expression leads to the following iteration,

$$\mathbf{x}^{k+1} \leftarrow \mathcal{F}_t{}^*\{\mathcal{S}_{\rho\alpha}(\mathcal{F}_t(\mathbf{x}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{x}^k) - \mathbf{y})))\}. \tag{5.5}$$

Based on this result, a fast proximal gradient (FPG) is derived in algorithm 5.1 that is referred to as *sparse signal recovery via fast proximal gradient* (S-FPG).

Algorithm 5.1 Sparse signal recovery via fast proximal gradient (S-FPG)

Input: $\mathbf{y}$, $\alpha \geq 0$, $\rho > 0$
Initialise: $k = 0$, $\mathbf{x}^0 = \mathbf{w}^0 = \mathbf{0}, t^0 = 1$
while stopping criterion is not met **do**
  $\mathbf{x}^{k+1} \leftarrow \mathcal{F}_t{}^*\{\mathcal{S}_{\rho\alpha}(\mathcal{F}_t(\mathbf{w}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{w}^k) - \mathbf{y})))\}$
  $t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2})$
  $\mathbf{w}^{k+1} \leftarrow \mathbf{x}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{x}^{k+1} - \mathbf{x}^k)$
end while
Output: $\widehat{\mathbf{x}} = \mathbf{x}^k$

5.2.3 Low-rank matrix recovery

In this section, we consider the use of low-rank as a regularisation prior. First, the Casorati matrix is formed so that each column represents a vectorised MR image of the sequence, see Eq. (3.59). This matrix is approximately low-rank with only a few significant singular values due to the correlation between images in time.

The finite-dimensional discrete model is

$$\mathbf{y} = \mathcal{E}(\mathbf{X}) + \mathbf{n}, \tag{5.6}$$

where in this case $\mathcal{E} : \mathbb{C}^{N_x N_y \times N_t} \to \mathbb{C}^M$ is the MRI encoding operator modelling both the random sub-Nyquist sampling and Fourier transform with $M \ll N_x N_y N_t$, and $\mathbf{X} \in \mathbb{C}^{N_x N_y \times N_t}$ represents the matrix to recover.

We consider the convex minimisation problem using the nuclear norm,

$$\mathbf{X}^\star = \arg\min_{\mathbf{X}\in\mathbb{C}^{N_x N_y \times N_t}} \Big\{ F(\mathbf{X}) \equiv \frac{1}{2}\|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha\|\mathbf{X}\|_* \Big\}. \tag{5.7}$$

The proximal mapping of the nuclear norm is known and has been described in section 4.2.3. Hence, this problem can be solved efficiently with the proximal gradient step

$$\mathbf{X}^{k+1} \leftarrow \text{prox}_{\rho\alpha\|.\|_*}(\mathbf{X}^k - \rho\nabla f(\mathbf{X}^k)). \tag{5.8}$$

Developing this expression leads to the following iteration,

$$\mathbf{X}^{k+1} \leftarrow \text{SVT}^{\mathcal{S}_{\rho\alpha}}(\mathbf{X}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{X}^k) - \mathbf{y})), \tag{5.9}$$

where $\text{SVT}^{\mathcal{S}}$ denotes the singular value soft thresholding operator defined as in Eq. (4.14). Based on the algorithm by Toh and Yun [137] which itself was inspired by FISTA from Beck and Teboulle [110], we propose to use the fast proximal gradient of the nuclear norm problem for solving the low-rank MRI reconstruction problem (algorithm 5.2). The algorithm is named *low-rank matrix recovery via fast proximal gradient* (LR-FPG).

Algorithm 5.2 Low-rank matrix recovery via fast proximal gradient (LR-FPG)

Input: $\mathbf{y}$, $\alpha \geq 0$, $\rho > 0$
Initialise: $k = 0$, $\mathbf{X}^0 = \mathbf{W}^0 = \mathbf{0}, t^0 = 1$
while stopping criterion is not met **do**
  $\mathbf{X}^{k+1} \leftarrow \text{SVT}^{\mathcal{S}_{\rho\alpha}}(\mathbf{W}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{W}^k) - \mathbf{y}))$
  $t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2})$
  $\mathbf{W}^{k+1} \leftarrow \mathbf{X}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{X}^{k+1} - \mathbf{X}^k)$
end while
Output: $\widehat{\mathbf{X}} = \mathbf{X}^k$

5.2.4 Simultaneously low-rank matrix and sparse signal recovery

We finally consider the combination of both low-rank and sparsity priors to evaluate the potential benefit of joint regularisation. We consider the following convex minimisation problem,

$$\mathbf{X}^\star = \arg\min_{\mathbf{X}\in\mathbb{C}^{N_x N_y \times N_t}} \Big\{ F(\mathbf{X}) \equiv \frac{1}{2}\|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha\|\mathbf{X}\|_* + \beta\|\mathcal{F}_t(\mathbf{X})\|_1 \Big\}, \tag{5.10}$$

where $f(\mathbf{X}) = \frac{1}{2}\|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2$ represents the data fidelity term, $g_1(\mathbf{X}) = \alpha\|\mathbf{X}\|_*$ is the low-rank prior and $g_2(\mathbf{X}) = \beta\|\mathcal{F}_t(\mathbf{X})\|_1$ is the sparsity prior with the temporal Fourier transform. This formulation generalises the previous problems since for $\alpha = 0$ it corresponds to the ℓ_1 norm regularised problem, and for $\beta = 0$, it corresponds to the nuclear norm regularised problem.

To minimise the objective function in (5.10), we propose to derive two algorithms. The first one is based on fast composite splitting and named *low-rank matrix and sparse signal recovery via fast composite splitting* (LRS-FCS). It is described in algo-

rithm 5.3. We also derive a method based on generalised forward-backward splitting (GFBS) (algorithm 5.4) that is referred to as *low-rank matrix and sparse signal recovery via generalised forward-backward splitting* (LRS-GFBS). Both methods were described previously in section 4.3.1 in a more general setting.

Algorithm 5.3 Low-rank matrix and sparse signal recovery via fast composite splitting (LRS-FCS)

Input: $\mathbf{y}$, $\alpha \geq 0$, $\beta \geq 0$, $\rho > 0$
Initialize: $k = 0$, $\mathbf{X}^0 = \mathbf{W}^0 = \mathbf{0}, t^0 = 1$
while stopping criterion is not met **do**
$\mathbf{H}_1^{k+1} \leftarrow \mathrm{SVT}^{\mathcal{S}_{2\rho\alpha}}(\mathbf{W}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{W}^k) - \mathbf{y}))$
$\mathbf{H}_2^{k+1} \leftarrow \mathcal{F}_t{}^*\{\mathcal{S}_{2\rho\beta}(\mathcal{F}_t(\mathbf{W}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{W}^k) - \mathbf{y})))\}$
$\mathbf{X}^{k+1} \leftarrow \frac{1}{2}(\mathbf{H}_1^{k+1} + \mathbf{H}_2^{k+1})$
$t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2})$
$\mathbf{W}^{k+1} \leftarrow \mathbf{X}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{X}^{k+1} - \mathbf{X}^k)$
end while
Output: $\widehat{\mathbf{X}} = \mathbf{X}^k$

Algorithm 5.4 Low-rank matrix and sparse signal recovery via generalised forward-backward splitting (LRS-GFBS)

Input: $\mathbf{y}$, $\alpha \geq 0$, $\beta \geq 0$, $\rho > 0$
Initialize: $k = 0$, $\mathbf{X}^0 = \mathbf{H}_1^0 = \mathbf{H}_2^0 = \mathbf{0}$
while stopping criterion is not met **do**
$\mathbf{H}_1^{k+1} \leftarrow \mathbf{H}_1^k + \mathrm{SVT}^{\mathcal{S}_{2\rho\alpha}}(2\mathbf{X}^k - \mathbf{H}_1^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{X}^k) - \mathbf{y}) - \mathbf{X}^k)$
$\mathbf{H}_2^{k+1} \leftarrow \mathbf{H}_2^k + \mathcal{F}_t{}^*\{\mathcal{S}_{2\rho\beta}(\mathcal{F}_t(2\mathbf{X}^k - \mathbf{H}_2^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{X}^k) - \mathbf{y}) - \mathbf{X}^k)\}$
$\mathbf{X}^{k+1} \leftarrow \frac{1}{2}(\mathbf{H}_1^{k+1} + \mathbf{H}_2^{k+1})$
end while
Output: $\widehat{\mathbf{X}} = \mathbf{X}^k$

5.2.5 Sampling considerations

Constraints

Notions of coherence, RIP or null space property are necessary to derive recovery guarantees in sparse signal and low-rank matrix recovery, as we have explained in section 3.3.3. However, these notions are not particularly convenient to design practical sampling strategies. Instead, random matrices or random sampling operators are used to implicitly achieve these properties. In particular, it has been shown that certain random matrices such as random Gaussian matrices satisfy the RIP with high probability.

However, sampling a truly random set of **k**-space samples is generally impractical in MRI because sampling patterns must also respect MR physics constraints [34]. A fundamental constraint is that gradient magnetic fields can only generate **k**-space trajectories that follow relatively continuous lines and curves. Consequently, the

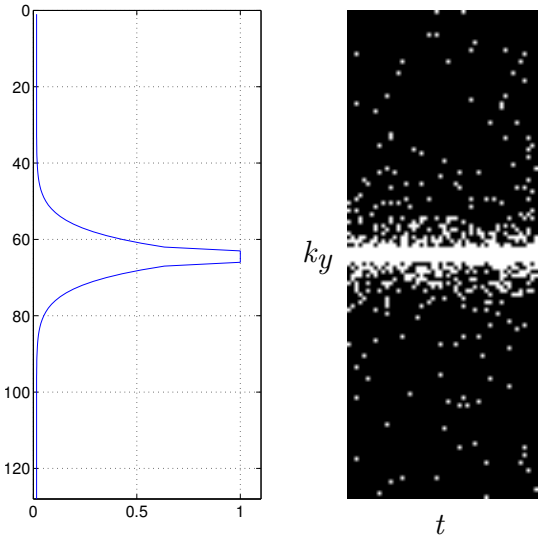

Figure 5.1: Polynomial variable density sampling pattern adapted for acquiring $(\mathbf{k}, t)$-space samples. Left figure shows the probability density function and right figure presents the (k_y,t) sampling pattern.

design of sampling patterns in MRI are generally not created with incoherence or RIP in mind, but rather with MR constraints and prior knowledge about the fact that MRI samples are acquired in the Fourier space. Even if incoherence property or RIP are not exactly satisfied in practice, it has been shown that signals could still be reconstructed with good fidelity.

Sampling schemes

Since a conventional strategy to acquire Fourier samples in MRI is along parallel equispaced **k**-space lines onto a Cartesian grid, a convenient way to achieve random undersampling is to randomly select fewer lines. However, since the energy distribution of MR images in **k**-space is known to be concentrated close to the center, a common strategy consists of densely sampling central **k**-space lines and randomly selecting lines elsewhere. Although selection of random lines can be drawn from a simple uniform probability distribution, a better approach is to give lower probabilities to the selection of lines nearer to the **k**-space edges in order to take into account the energy distribution and also because it may overcome coherence problems at low spatial frequencies for some sparsifying transforms [92]. This sampling strategy is often referred to as *polynomial variable density* (PVD) sampling [33]. Here, a similar sampling strategy is adapted to dynamic imaging, where the PVD sampling is applied for each acquisition time frame as shown in figure 5.1.

In this chapter, we will also use non-Cartesian sampling schemes. We will use pseudo-radial sampling with 2D projections at the same angle (equispaced) and 2D projections at different angles based on the golden angle ratio. In contrast to

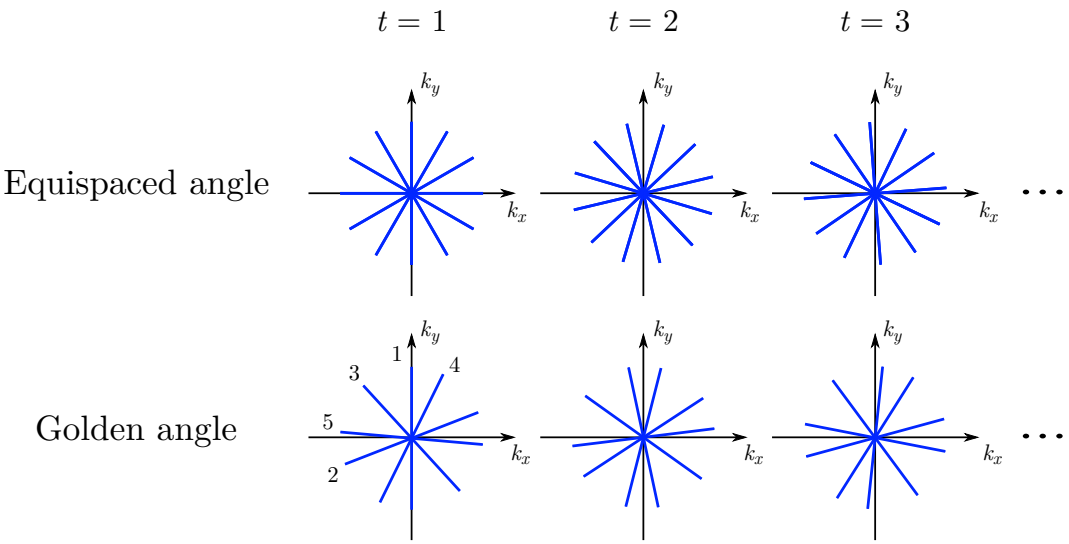

Figure 5.2: Equispaced and golden angle radial sampling with random rotations of the whole pattern for each time frame.

equispaced angle radial lines, the golden angle method selects a projection based on the golden angle, where a new projection is determined from the previous one by applying an anticlockwise rotation of the golden angle, equal to $180°/\frac{1+\sqrt{5}}{2} \approx 111.25°$. Golden angle radial sampling has been shown to be appealing in dynamic MRI [138] and to demonstrate good results from a compressive sensing perspective [139]. The latter can be justified by the fact that higher spatial and temporal incoherence may be achieved due to unequally spaced angles. Since we deal with dynamic imaging, a form of randomness must also be present in time. We employ a simple strategy with random rotations of the whole sampling pattern between each frame. Figure 5.2 shows equispaced and golden angle radial sampling schemes.

Note that these undersampling strategies can be achieved easily by omitting readouts from conventional Cartesian or radial acquisitions, which makes them particularly suitable and inexpensive from an MR acquisition point of view since only minor pulse sequence modifications are required. In this thesis however, datasets are either directly created in the image domain (for simulated data) or obtained from an MRI scanner using standard imaging sequences at the Nyquist rate (for *in vivo* data). Datasets are then retrospectively undersampled to evaluate reconstruction methods. This offers two major advantages: a *ground truth* is accessible for comparison, and multiple sampling schemes can be investigated without requiring a new acquisition on a real MR scanner.

5.3 Numerical simulations

5.3.1 Framework

We use the following numerical phantoms and *in vivo* cardiac datasets in this work:

- **Breath-hold Shepp-Logan based numerical phantom ("SL phantom")**. A numerical phantom based on the Shepp-Logan phantom (dimensions $N_x = N_y = 128$, $N_t = 80$). This phantom includes local intensity changes and periodic local motion in magnitude images and a spatially-smooth phase that slowly varies in time randomly. This phantom is referred to later on by "SL phantom".

- **Free-breathing numerical phantom ("PINCAT phantom")**. The free-breathing *physiologically improved non-uniform cardiac torso* (PINCAT) phantom (dimensions $N_x = N_y = 128$, $N_t = 50$). The PINCAT phantom is an adaptation to MR of the NCAT phantom in CT and was proposed by Sharif and Bresler [140] for evaluating cardiac MR imaging schemes and reconstruction methods in MRI. The version we use has been modified by Lingala et al. [82] to include intensity changes (simulating perfusion dynamics) and respiratory motion. We further modified it to produce a complex-valued signal by adding a spatially-smooth phase that slowly varies in time. This phantom is referred to later on by "PINCAT phantom".

- **Breath-hold cardiac**. A complex-valued breath-hold cardiac dataset (dimensions $N_x = 128$, $N_y = 128$, $N_t = 50$) which was acquired with a steady-state free precession sequence at Nyquist rate with a Cartesian sampling scheme on a Philips 1.5T MRI scanner.

- **Free-breathing cardiac**. A free-breathing cardiac dataset (dimensions $N_x = N_y = 128$, $N_t = 90$) from a 3T MRI scanner. This dataset is based on noisy magnitude-reconstructed images from an MR scanner and as such was initially real-valued without any phase information available. As with the numerical phantoms, a spatially-smooth phase that slowly varies in time is added post-acquisition.

The spatially-smooth and slowly time-varying phase added on some of these datasets is generated for each time frame independently and was adapted from Ref. [141]. It is illustrated in figure 5.3. The reason to add a time-varying random phase is to obtain more realistic experiments by simulating complex-valued signal. Intensity of all datasets were normalised between values 0 and 255 in magnitude prior to any processing. Gaussian noise was added explicitly on each real and imaginary channel with a standard deviation $\sigma = 5$, except for the breath-hold cardiac dataset

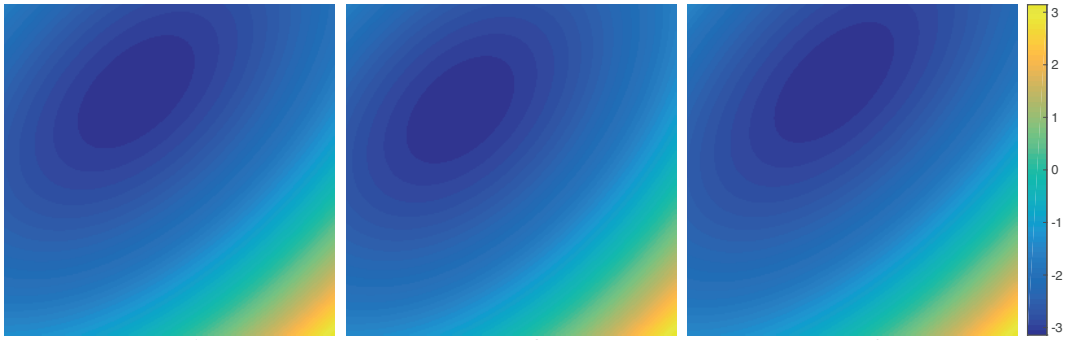

Figure 5.3: Spatially-smooth and slowly time-varying phase generated for three time frames.

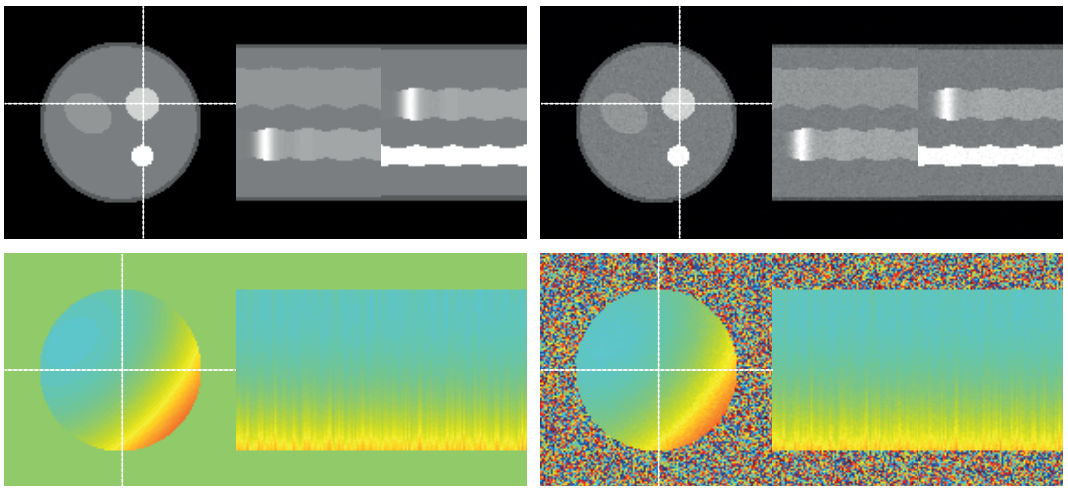

Figure 5.4: SL phantom showing magnitude (top) and phase (bottom) time frames as well as associated x-t and y-t temporal profiles along the dashed lines. Left: noiseless, right: noisy.

which already contained noise in both real and imaginary channels. For the free-breathing cardiac dataset, supplementary noise was added in both channels even if some noise was originally present in magnitude-reconstructed images. Note that for *in vivo* datasets, the ground truth without noise is not known, hence in this case the ground truth will refer to the noisy signal. The numerical phantoms are shown in figures 5.4 (SL phantom) and 5.5 (PINCAT phantom). Cardiac datasets are shown in figures 5.6 (breath-hold) and 5.7 (free-breathing).

The motivations in testing these four datasets are to characterise and evaluate the performance of reconstruction methods in different states of dynamic imaging: (i) phantom signals with and without global motion, (ii) real data with and without global motion. The presence of the global (respiratory-like) motion will generally deteriorate the reconstruction quality and as such it is interesting to see how methods perform in more difficult conditions. Similarly, reconstruction quality can vary depending on the nature of the input signal, i.e. numerical phantoms versus real

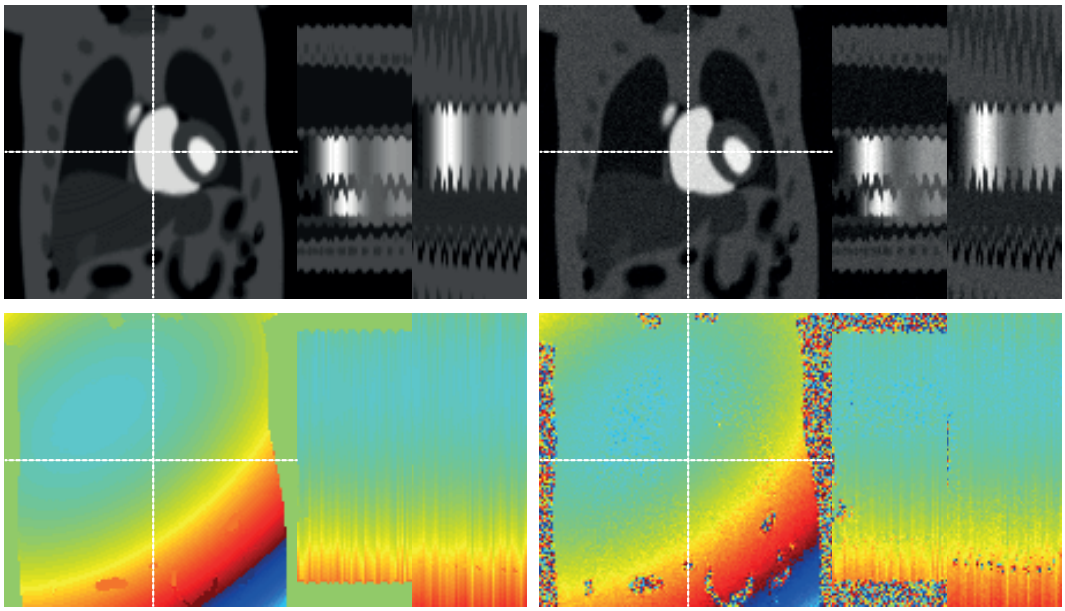

Figure 5.5: PINCAT phantom showing magnitude (top) and phase (bottom) time frames as well as associated x-t and y-t temporal profiles along the dashed lines. Left: noiseless, right: noisy.

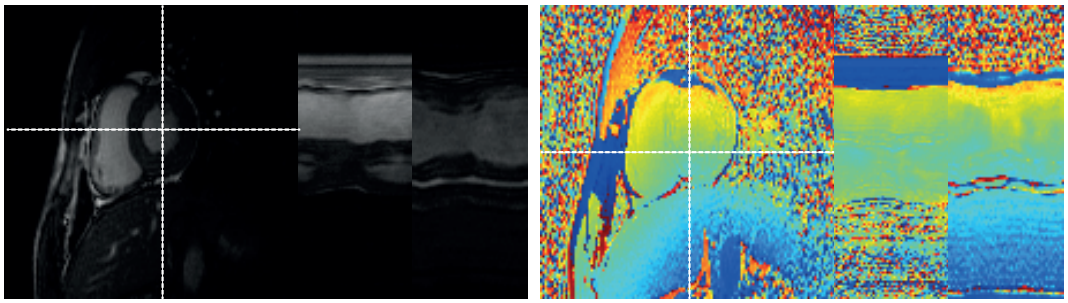

Figure 5.6: Breath-hold cardiac dataset showing magnitude (left) and phase (right) time frames as well as associated x-t and y-t temporal profiles along the dashed lines.

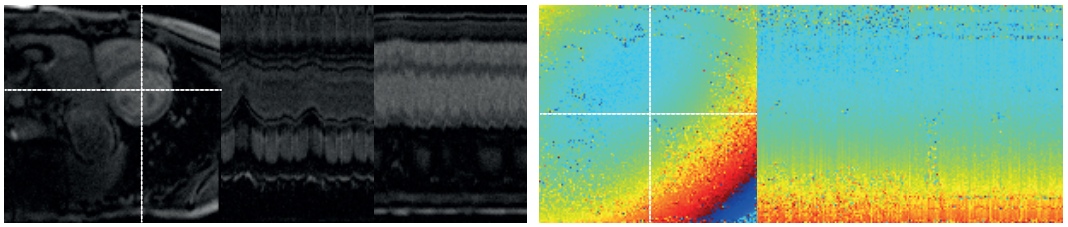

Figure 5.7: Free-breathing cardiac showing magnitude (left) and phase (right) time frames as well as associated x-t and y-t temporal profiles along the dashed lines.

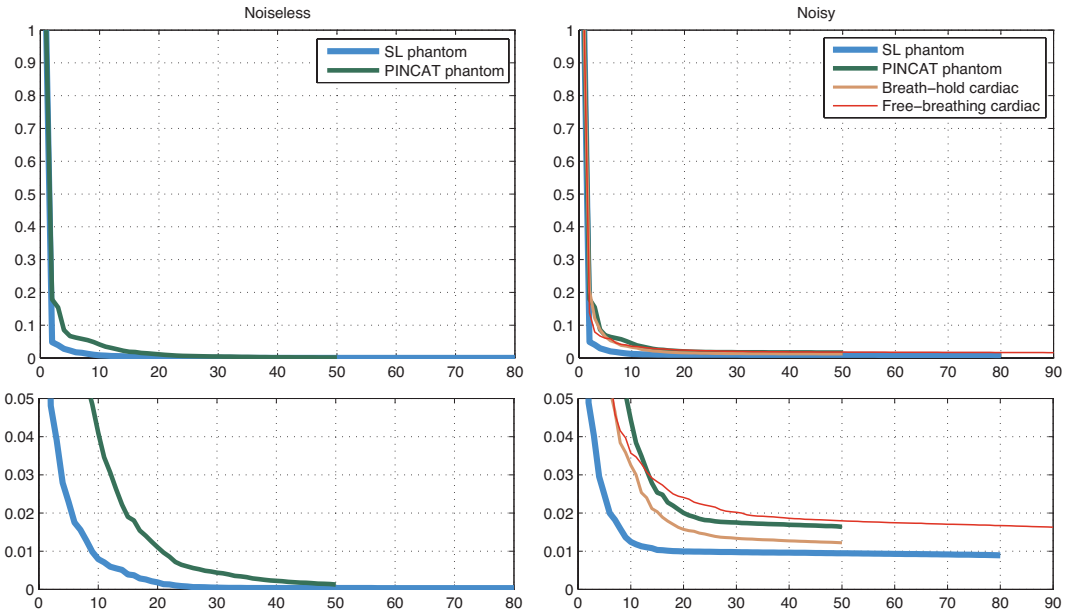

Figure 5.8: Singular values (normalised) for the dynamic MRI datasets with zoom-in graphs (bottom).

	$\|.\|_1$	$\|\mathcal{F}_t(.)\|_1$	Decrease (%)
SL phantom	7.18e+07	1.79e+07	25.0
PINCAT phantom	4.33e+07	1.62e+07	37.4
Breath-hold cardiac	1.45e+07	4.97e+06	34.2
Free-breathing cardiac	5.71e+07	2.01e+07	35.2

Table 5.1: Sparsity characterisation of the different datasets. Right column represents the decreasing amount of sparsity in percentage, e.g. the PINCAT phantom benefits most from the temporal Fourier transform.

data. Datasets are characterised by their singular values and their sparseness respectively in figure 5.8 and table 5.1 to give some intuition about their sparsity levels and ranks. We also show intensity profiles in figure 5.9.

Numerical simulations are performed in Matlab on a Linux platform. We refer the reader to appendix B for the error metrics used to evaluate the performance of the reconstructions. For reference, we compute the zero-filled inverse Fourier transform (ZF-IDFT) and a very basic sliding window reconstruction (zeroth-order). We also compare our method with three state of the art methods, **k**-t FOCUSS with temporal average [73–75] and two low-rank and sparse techniques, **k**-t SLR [82] and PS-Sparse [86]. These methods were described in section 3.4. **k**-t FOCUSS is implemented with a maximum of 40 inner iterations (conjugate gradient step), 2 outer iterations (FOCUSS step) and weighting matrix power factor of $p = 0.5$ (cf. section 3.4.2). The low-resolution initial estimate is obtained by using a zero-filled inverse Fourier transform using the low-frequency samples (figure 5.10). **k**-t SLR is implemented with penalty parameters set to $\beta_1 = \beta_2 = 10^{-7}$ for Schatten and TV norms and

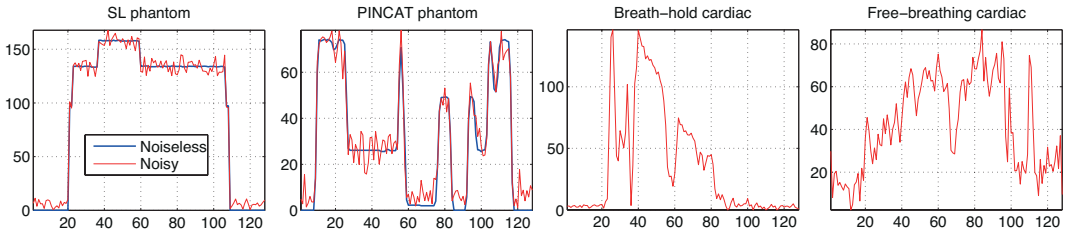

Figure 5.9: Intensity profiles along lines taken in the middle of the images.

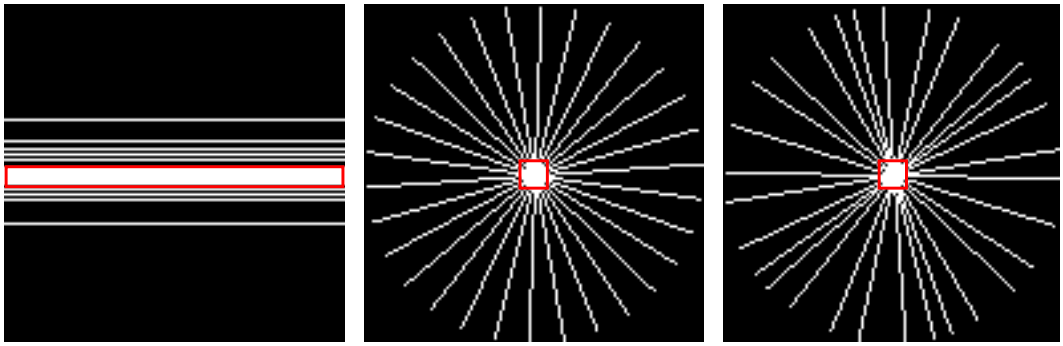

Figure 5.10: Sampling patterns used in this study, showing here only one acquisition time frame. From left to right, PVD (Cartesian sampling), equispaced and golden angle radial sampling schemes. The red contours show the fully (or almost fully) sampled $(\mathbf{k}, t)$-space sets that are used in $\mathbf{k}$-t FOCUSS and PS-Sparse respectively to obtain the low-resolution estimate and to evaluate the basis for the temporal subspace.

their respective incrementation parameters are set to 25. A maximum number of 50 inner and 9 outer iterations is chosen. PS-Sparse is implemented with the maximum number of iterations of 5 for outer loop and 100 for inner loop, the tolerance is set to 10^{-5} and the initial value of continuation parameter to 1000. Note that initially PS-Sparse has been designed with random Cartesian undersampling in mind with a subdivision of the sampling pattern into two sets, where one set corresponds to the fully sampled central lines of the $\mathbf{k}$-space to estimate the basis for the temporal subspace. We have adapted PS-Sparse to handle radial sampling trajectories so that the fully sampled $(\mathbf{k}, t)$-space part corresponds to the fully sampled centre of the radial sampling pattern. This is illustrated in figure 5.10.

For the numerical simulations, different sampling schemes described in section 5.2.5 are tested (PVD, equispaced angle radial and golden angle radial) with an acceleration factor of about 10, which means that only about 10% of (random) Fourier samples are acquired. In this context, note that reconstructions are particularly challenging because we are dealing with complex-valued, noisy, motion-included and undersampled data. Note that most of the time reconstruction techniques are rarely evaluated with all these characteristics together.

The regularisation parameter for each algorithm is tested for a range of different values and the best one is selected in terms of reconstruction quality according to

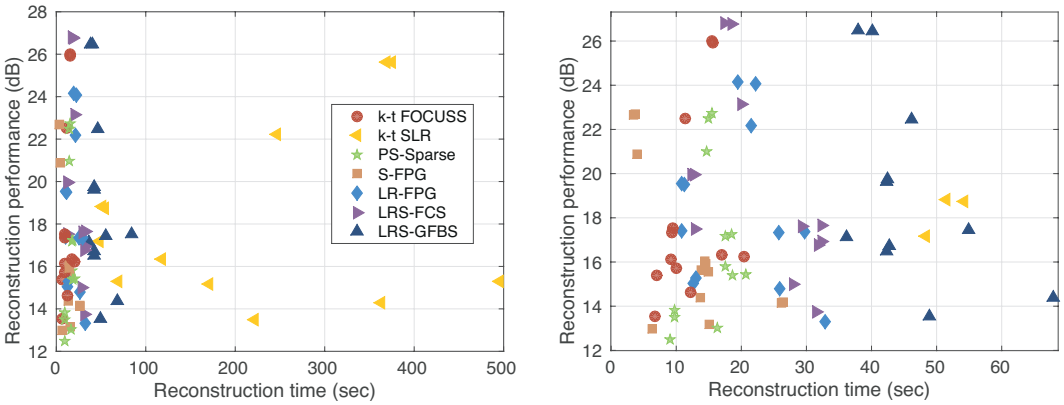

Figure 5.11: Comparison of the different methods in terms of reconstruction performance versus reconstruction time. Each point represents a computed reconstruction from table 5.2. Right figure is a close-up of the left figure.

Eq. (B.2). The set of regularisation parameters tested is

$$\{0, 10^k : k \in \mathbb{Z}; -4 \leq k \leq 3\}. \tag{5.11}$$

For multiple priors, the best combination of regularisation parameters is selected. For PS-Sparse (see Eq. (3.63)), the set of regularisation parameters tested is different because the PS model order directly relates to the rank of the reconstructed data. The PS model order range tested is $\{0, 2, 4, 8, 16, 32\}$.

The step sizes for the proximal gradient methods are chosen as $\rho = 1/L$ with $L = 1$ according to appendix A.3. The stopping criterion for the proximal gradient algorithms is defined as follows. Algorithms are stopped if a maximum number of 100 iterations is reached, or if the corresponding objective function F does not decrease significantly any more, i.e. when

$$\frac{F(\mathbf{x}^{k+1}) - F(\mathbf{x}^k)}{F(\mathbf{x}^k)} \leq \texttt{tol}, \tag{5.12}$$

where `tol` is the tolerance set to 10^{-5}.

5.3.2 Quantitative reconstruction results

Reconstruction results are reported in table 5.2 and should be interpreted in decibels as they have been computed with Eq. (B.2). To complement this table, we also show

- in figure 5.11, the reconstruction times needed for algorithms to achieve the reconstruction results of table 5.2,
- in figure 5.12, the reconstruction performance as a function of the computational time for each algorithm in the case of SL phantom and PVD sampling,

SL phantom	PVD	Equi. angle	Golden angle
ZF-IDFT	14.2	15.7	15.2
Sliding window	19.5	20.2	20.0
k-t FOCUSS [73–75]	22.5	26.0	25.9
k-t SLR [82]	22.2	25.6	25.6
PS-Sparse [86]	21.0	22.7	22.5
S-FPG	20.9	22.7	22.7
LR-FPG	22.2	24.1	24.2
LRS-FCS	**23.1**	**26.8**	**26.8**
LRS-GFBS	22.5	26.4	26.5

(a) SL phantom.

PINCAT phantom	PVD	Equi. angle	Golden angle
ZF-IDFT	12.0	12.6	12.3
Sliding window	13.2	13.5	13.2
k-t FOCUSS [73–75]	15.4	17.5	17.3
k-t SLR [82]	17.2	18.7	18.8
PS-Sparse [86]	15.8	17.2	17.2
S-FPG	14.4	15.6	15.6
LR-FPG	17.4	19.5	19.6
LRS-FCS	**17.5**	**20.0**	**20.0**
LRS-GFBS	17.1	19.6	19.8

(b) PINCAT phantom.

Breath-hold cardiac	PVD	Equi. angle	Golden angle
ZF-IDFT	8.5	7.0	6.6
Sliding window	11.8	11.8	11.7
k-t FOCUSS [73–75]	13.5	16.1	15.7
k-t SLR [82]	13.5	15.3	15.2
PS-Sparse [86]	12.5	13.8	13.5
S-FPG	13.2	16.1	15.9
LR-FPG	13.3	15.3	15.0
LRS-FCS	**13.7**	**16.9**	**16.8**
LRS-GFBS	13.5	16.7	16.5

(c) Breath-hold cardiac.

Free-breathing cardiac	PVD	Equi. angle	Golden angle
ZF-IDFT	8.6	10.1	9.8
Sliding window	12.0	13.2	13.0
k-t FOCUSS [73–75]	14.6	16.3	16.2
k-t SLR [82]	14.3	15.3	16.4
PS-Sparse [86]	13.0	15.4	15.4
S-FPG	13.0	14.1	14.2
LR-FPG	14.8	17.4	17.4
LRS-FCS	**15.0**	**17.6**	**17.6**
LRS-GFBS	14.4	17.5	17.5

(d) Free-breathing cardiac.

Table 5.2: Quantitative reconstruction results (in dB).

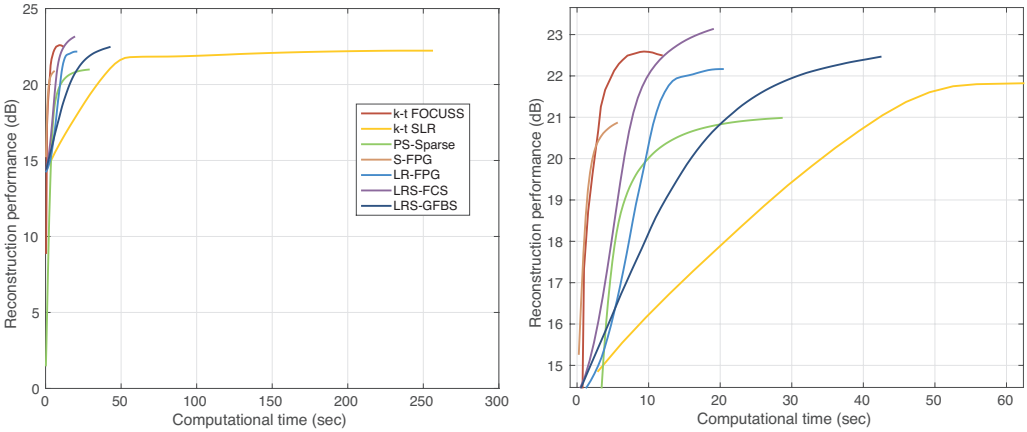

Figure 5.12: Comparison of the different methods in term of reconstruction performance versus computational time for SL phantom and PVD sampling. Right figure is a close-up of the left figure.

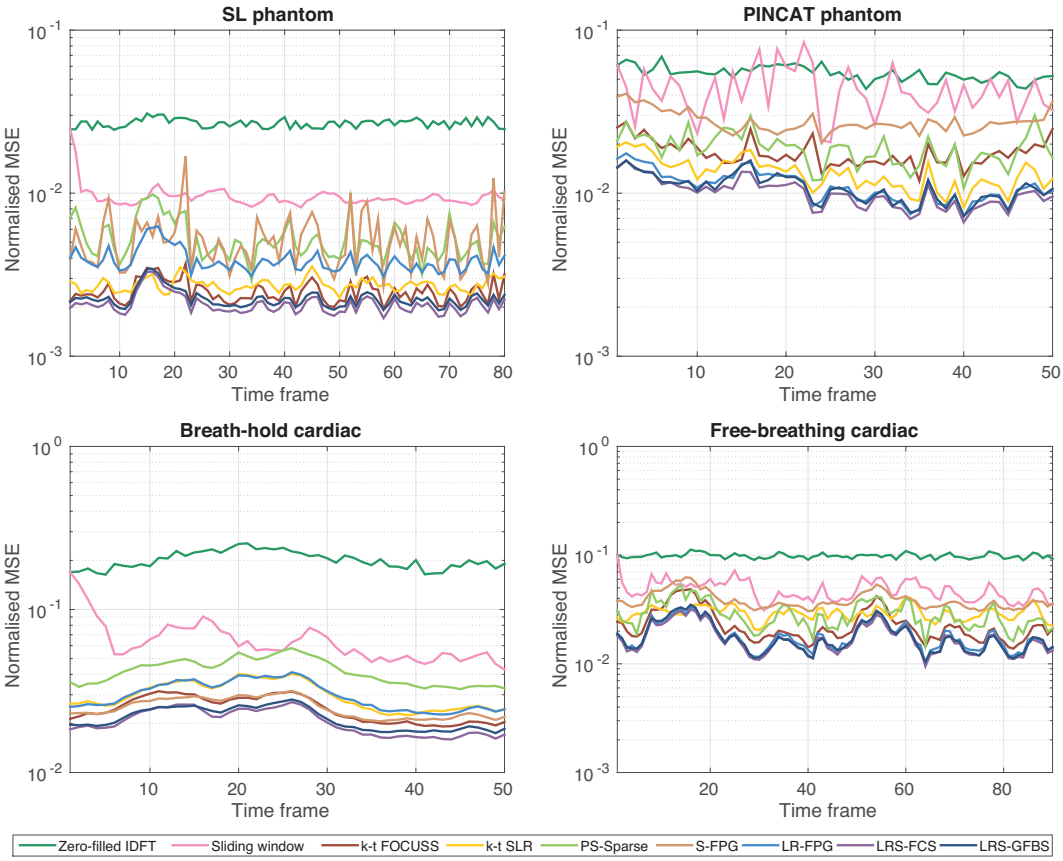

Figure 5.13: Normalised mean square error at each time frame for the different datasets using equispaced angle radial sampling.

- in figure 5.13, the normalised mean square error at each time frame as in Eq. (B.3) in the case of equispaced angle radial sampling to obtain more insight of the reconstruction performance of each methods per time frame.

From these results, first note that as expected, sliding window reconstructions provide improved reconstruction quality results over standard inverse Fourier reconstructions (ZF-IDFT). However, while the overall reconstruction performances of sliding window are higher than ZF-IDFT reconstructions, the sliding window reconstruction obtains inconsistent errors in time and presents important error differences from one time frame to the next one as observed in figure 5.13 for the PINCAT phantom. This can be explained by the fact that the PINCAT phantom presents considerable motion in between time frames, and because our sliding window reconstruction is of zeroth-order (the missing **k**-space points are estimated based only on the previous **k**-space time frame).

In these experiments, PS-Sparse did not show any major advantages as reconstruction results were often of lower quality than other state of the art methods (**k**-t FOCUSS, **k**-t SLR). **k**-t SLR provided superior results compared to PS-Sparse in almost every cases but surprisingly **k**-t FOCUSS provided consistently very close or better results than **k**-t SLR in 3 out of 4 datasets tested. Perhaps more importantly, **k**-t SLR was shown to suffer from much longer computational times compared to other methods due to the optimisation scheme employed.

These simulations indicate that LRS-FCS systematically produces higher reconstruction quality results than all other methods, in very competitive computational times. Figure 5.12 shows that LRS-FCS converges faster to a solution of higher quality than other methods. In fact, figures 5.11 and 5.12 make it clear that fast proximal gradient methods (S-FPG, LR-FPG, LRS-FCS) can rival with **k**-t FOCUSS, which is arguably one of the fastest state of the art method in dynamic MR reconstruction.

Comparing the two proximal gradient algorithms combining low-rank and sparsity prior (LRS-FCS, LRS-GFBS), it seems to indicate that LRS-FCS is consistently faster while obtaining slightly better reconstruction results. These two algorithms minimise the same functional (5.10) and are not drastically different. LRS-GFBS is mathematically more rigorous because the actual convergence of the sequence has been proved. LRS-FCS is shown to converge faster because of the accelerating scheme borrowed from FISTA as shown clearly in figure 5.12. LRS-GFBS still provides enhanced reconstruction quality results than most other methods but since it takes longer computational times than LRS-FCS, we will not discuss this algorithm hereinafter.

Interestingly, it should be noted that reconstructions using the convex nuclear norm as regularisation (LR-FPG) almost consistently yield superior reconstruction results than using the convex ℓ_1 norm using the temporal Fourier domain as sparsi-

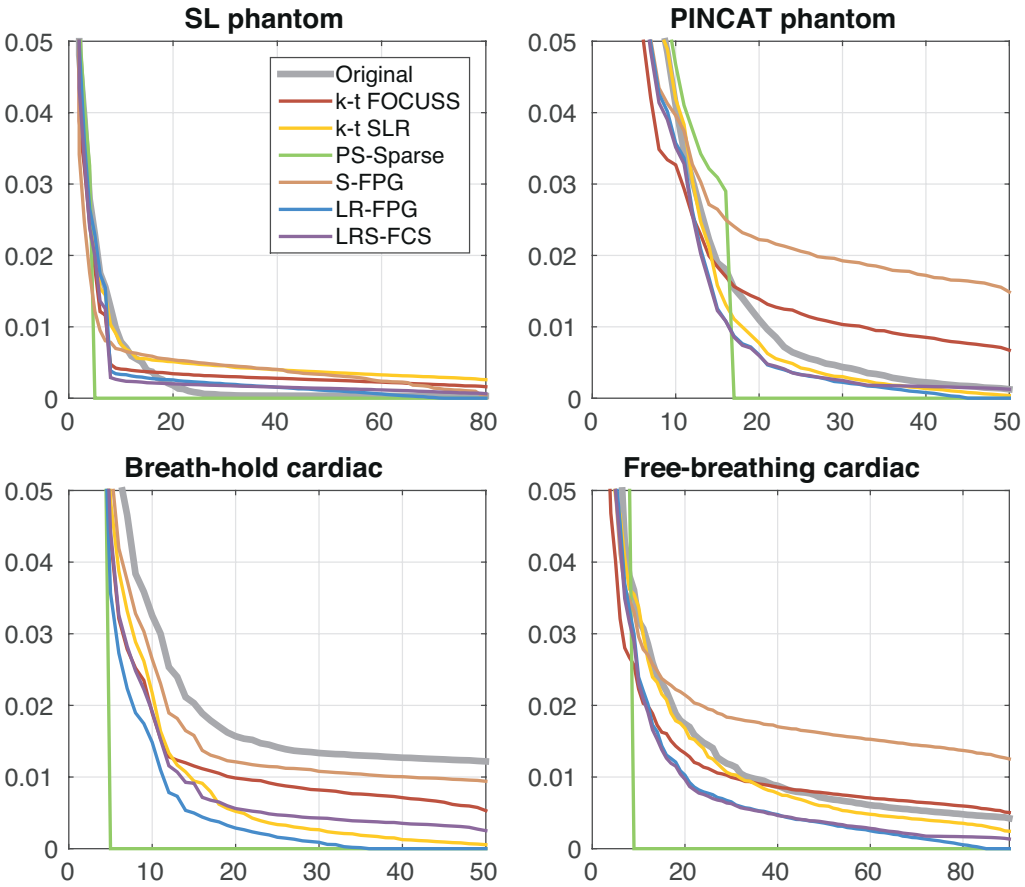

Figure 5.14: Singular values (normalised, zoom-in) for the different methods using equispaced radial sampling.

fying transform (S-FPG). This seems to indicate that the convex low-rank constraint itself is a serious alternative to standard compressed sensing techniques using convex sparsity constraint for dynamic MR imaging. This was one of our conclusion in our conference paper [135].

Finally, we show in figure 5.14 the Casorati matrix singular values of the various reconstruction methods for the different datasets in the case of equispaced angle radial sampling. Although singular values on their own are not an indicator of a good reconstruction, this figure mainly shows the difference between PS-sparse that severely enforces the matrix rank through its partial separable model order selection, and other methods that result in "naturally" closer singular values to the original signal.

5.3.3 Qualitative reconstruction results

We first present some qualitative results for the SL phantom with the golden angle radial sampling in figure 5.15. This figure does not show reconstructed images but error images in order to better evaluate the different reconstruction methods. Error

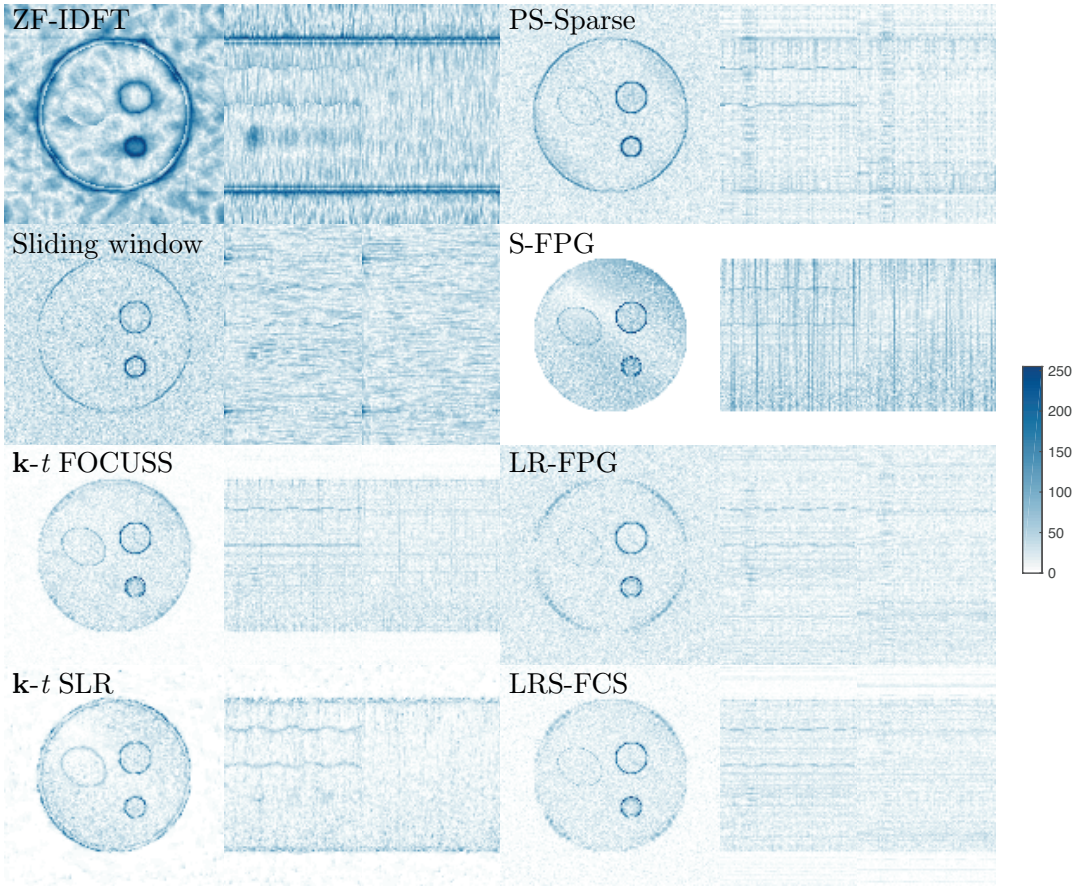

Figure 5.15: Error (×5) magnitude images for the SL phantom in the case of golden angle radial sampling. One frame and x-t and y-t profiles are shown for each method.

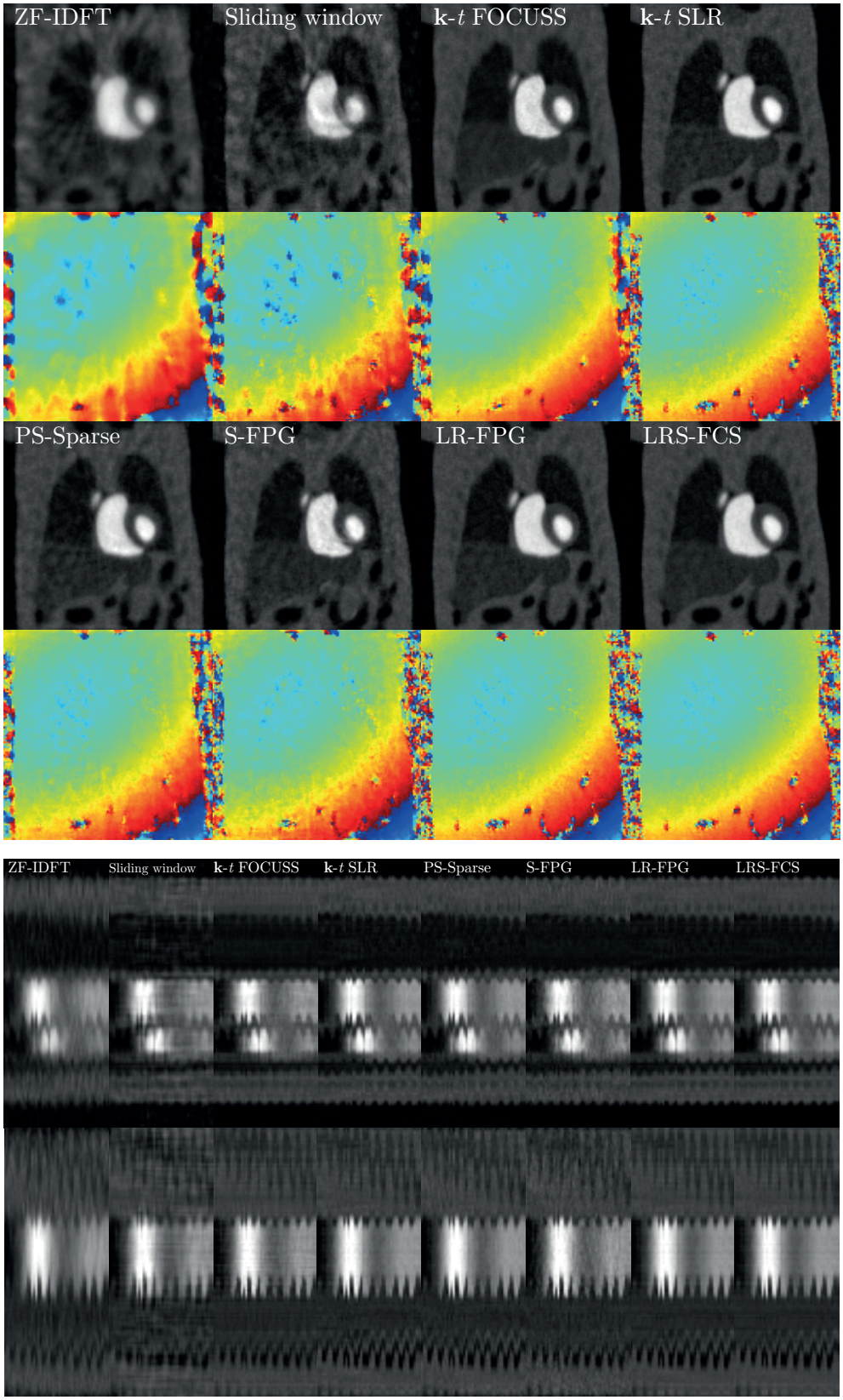

Figure 5.16: Reconstruction of the PINCAT phantom with equispaced angle sampling. Top figures show extracted magnitude and phase images from the sequence. Bottom figures show x-t and y-t profiles.

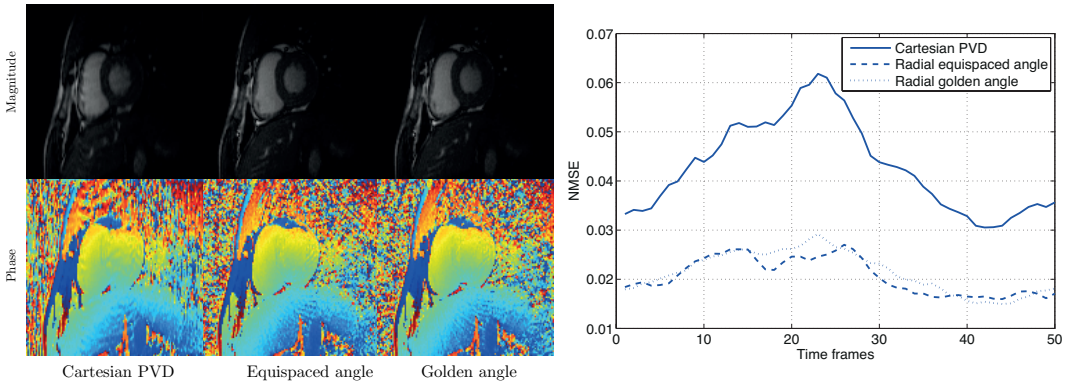

Figure 5.17: The influence of the sampling patterns with breath-hold cardiac reconstructions and LRS-FCS algorithm. Left figures show magnitude and phase images extracted from the reconstructed sequences, right figure shows the NMSE.

images (multiplied by a factor to amplify the differences) are simply computed by taking the absolute value of the difference between ground truth images and reconstructed images. Figure 5.15 supports numerical results obtained in the previous section, where good reconstruction results were obtained by **k**-t FOCUSS, **k**-t SLR and LRS-FCS. Finally, observe how S-FPG provides a denoising where no signal but only noise is present, while LR-FPG minimises the error more globally. The combination of S-FPG and LR-FPG results in LRS-FPG that takes the best of both worlds.

In figure 5.16, we show magnitude and phase images reconstruction of the PINCAT phantom with equispaced radial sampling. We also show image time profiles of the reconstructions. These images are to be compared with the ground truth images that were shown in figure 5.5. Although it might be difficult to see in the first instance, a careful inspection of both magnitude and phase images generally reveals that LRS-FPG provides better reconstruction in small details (e.g. less noisy and blurry) compared to other methods.

5.3.4 Radial sampling

Table 5.2 and figure 5.17 reveals that radial-based sampling schemes offer higher reconstruction results than Cartesian sampling, both quantitatively and from visual inspection. For example in figure 5.17 Cartesian random sampling shows blurring artefacts in the magnitude reconstructed image. This may be attributed to the fact that radial-based patterns sample more closely to the energy distribution in the Fourier space as most of the energy is situated in the centre of the **k**-space. While we could intuitively expect better results from the golden angle scheme due to potentially more "incoherency" provided by a non equispaced angle sampling, it is actually difficult to draw a definitive conclusion on the best choice between equispaced and golden angle radial patterns.

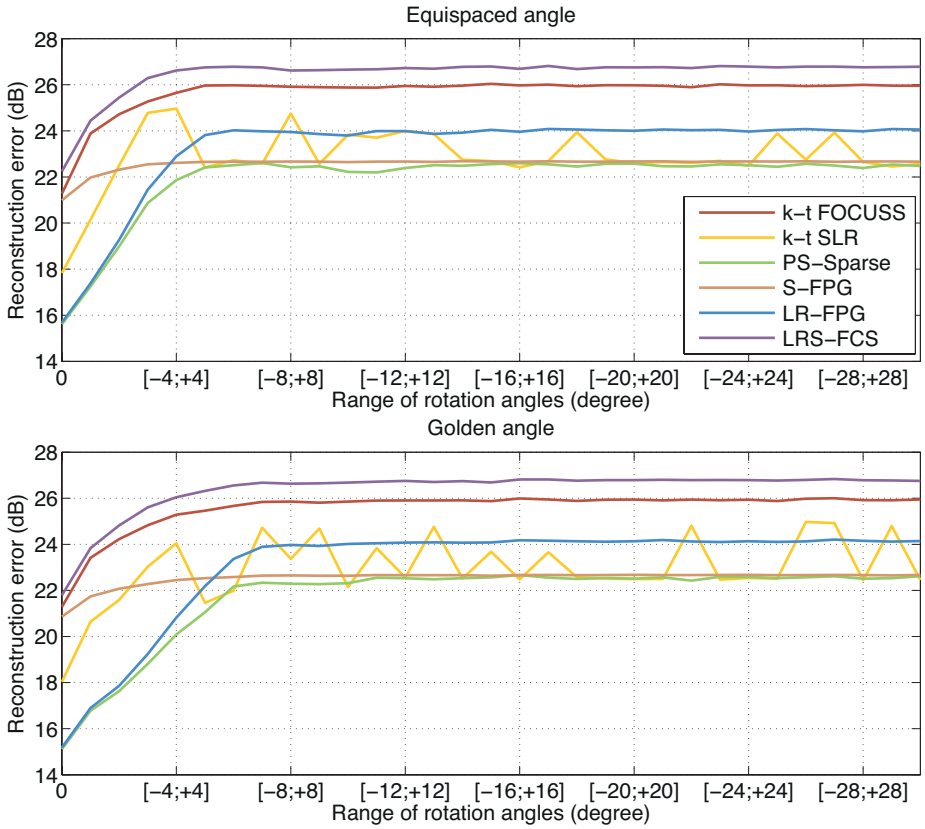

Figure 5.18: Influence of randomness in time for radial-based sampling schemes.

In this study, random rotation angles were chosen uniformly between $[-20°; +20°]$ to rotate radial sampling patterns across each acquisition frames. The influence of this random rotation angle on the reconstruction quality is an interesting question. In figure 5.18, we show the reconstruction results for various ranges of rotation angles using the SL phantom dataset. More specifically, reconstructions were computed for the different methods from no rotation (i.e. $0°$ which means no randomness and the sampling trajectory being consistent in time) to rotation angles chosen uniformly in the range $[-30°; +30°]$. This figure suggests there do not seem to be a range of rotation angles for which the reconstructions are higher: as long as the range is chosen equal or superior to about $[-6°; +6°]$, methods perform similarly (although for **k**-t SLR, this is more contestable as it slightly varies over the different ranges). This figure also seems to indicate that the influence of randomness in time is more important for methods that are based on low-rank and the Casorati matrix formulation than sparsity-based only methods. Indeed, the reconstruction result for S-FPG with no rotation (i.e. $0°$) is 21dB and 22.7dB when rotation angles are superior to about $[-6°; +6°]$. For LR-FPG, the reconstruction result is 15.7dB at $0°$ but 24dB when rotation angles are superior to $[-6°; +6°]$. The randomness of the sampling pattern seems to play a more important role in the latter case.

5.3.5 Nonconvex and hard thresholding approaches

As briefly discussed in section 4.5, nonconvex approaches have been shown to obtain higher results in CS. In this section, we examine nonconvex and greedy approaches against convex low-rank regularisation.

For the convex case, we use the low-rank regularisation method based on fast proximal gradient (algorithm 5.2). This method is based on the nuclear norm and the singular value soft thresholding operator $\text{SVT}^{\mathcal{S}}$ which consists in soft thresholding the singular values.

For the nonconvex case, we are interested in the nonconvex problem

$$\mathbf{X}^\star = \arg\min_{\mathbf{X}\in\mathbb{C}^{N_x N_y \times N_t}} \Big\{ F(\mathbf{X}) \equiv \frac{1}{2}\|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha\|\mathbf{X}\|_p \Big\}, \tag{5.13}$$

for $0 < p < 1$ where $\|.\|_p$ refers to the Schatten p-norm as defined in Eq. (A.8) (although for $p < 1$ this is not a norm anymore). The generalisation of the soft thresholding (or shrinkage) operator for various p was proposed by Chartrand in Refs. [84, 85] based on the generalisation of the Huber function. We denote it $\mathcal{S}_\alpha^p$ and it is defined element-wise as,

$$\mathcal{S}_\alpha^p(\mathbf{z}) = \{\mathcal{S}_\alpha^p(z_n)\}_{n=1}^N = \text{sgn}(z_n)(|z_n| - \alpha|z_n|^{p-1})_+. \tag{5.14}$$

Note that for $p = 1$, this is equivalent to $\mathcal{S}_\alpha$ as defined in Eq. (4.11). Following

	SL phantom	PINCAT	Cardiac (BH)	Cardiac (FB)
ZF-IDFT	16.7	13.6	7.6	10.8
Convex	24.1	**20.2**	**15.1**	**17.9**
Nonconvex, $p = 0.1$	22.1	19.1	11.9	14.8
Nonconvex, $p = 0.5$	24.3	19.7	14.8	17.4
Nonconvex, $p = 0.9$	**24.7**	19.7	13.8	17.6
Hard thresholding	21.0	14.1	10.5	14.3

Table 5.3: Reconstruction results (in dB) using Eq. (B.2) for low-rank regularisation via convex, nonconvex and hard thresholding approaches. ZF-IDFT reconstructions are reported for reference.

the above definition, the generalised singular value soft thresholding operator can be expressed as

$$\mathrm{SVT}^{\mathcal{S}_\alpha^p}(\mathbf{Z}) = \mathbf{U}\mathcal{S}_\alpha^p(\mathrm{diag}(\mathbf{\Sigma}))\mathbf{V}^{\mathrm{H}}. \tag{5.15}$$

where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ represents the singular value decomposition of $\mathbf{Z}$. We outline a simple algorithm to solve (5.13) with the following iteration,

$$\mathbf{X}^{k+1} \leftarrow \mathrm{SVT}^{\mathcal{S}_{\rho\alpha}^p}(\mathbf{X}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{X}^k) - \mathbf{y})). \tag{5.16}$$

The step size is chosen as $\rho = 1$ and the algorithm is stopped if a maximum number of 100 iterations is reached, or if Eq. (5.12) is satisfied.

Finally for the greedy approach, a simple iterative hard thresholding (IHT) algorithm named singular value projection as seen in Eq. (4.54),

$$\mathbf{X}^{k+1} \leftarrow \mathrm{SVT}^{\mathcal{H}_{\rho\alpha}}(\mathbf{X}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{X}^k) - \mathbf{y})), \tag{5.17}$$

where $\mathrm{SVT}^{\mathcal{H}}$ is the hard thresholding operator defined as in Eq. (4.53) and the step size is chosen as $\rho = 1$. The algorithm is stopped if a maximum number of 100 iterations is reached, or if the estimated matrix does not decrease significantly any more compared to the previous iteration, i.e. when $(\mathbf{X}^{k+1} - \mathbf{X}^k)/\mathbf{X}^k \leq 10^{-5}$.

For this experiment, we use 8-fold acceleration and equispaced radial sampling. Regularisation parameters were tested for the set defined as in (5.11) but extended to larger values after preliminary simulations with nonconvex and greedy approaches. Table 5.3 reports the reconstruction results for the three methods and the different datasets. This table shows that in general, the convex approach performs better over other methods. Interestingly, these results also seem to indicate that nonconvex optimisation offers higher reconstruction results as $p \to 1$, that is as convexity is approached. The IHT algorithm can be shown to improve a standard inversion using a zero-filled Fourier transform, but it does not offer superior reconstruction results compared to other methods.

5.3.6 Local low-rank matrix recovery

Another possibility for low-rank matrix recovery is to impose low-rank *locally* instead of globally. This was suggested by Trzasko and Manduca [142] in the context of dynamic MRI denoising. The idea is to exploit the fact that parts of the image sequence are more likely to be stronger rank-deficient locally than globally. This is particularly true for breath-hold sequences that can feature motionless subparts, and which are expected to be rank-1 or approximately rank-1. Locally low-rank can also remediate problems appearing when only a small number of temporal frames is considered. Indeed, if this is the case, the rank of the Casorati matrix $N_x N_y \times N_t$ may not be sufficiently small to consider the matrix as low-rank since $N_t \ll N_x N_y$.

This approach is studied here for undersampled dynamic MRI reconstruction by solving a slightly modified version of optimisation problem (5.7),

$$\mathbf{X}^\star = \arg \min_{\mathbf{X} \in \mathbb{C}^{N_x N_y \times N_t}} \Big\{ F(\mathbf{X}) \equiv \frac{1}{2} \|\mathcal{E}(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha \sum_{\omega \in \Omega} \|\mathcal{R}_\omega(\mathbf{X})\|_* \Big\}, \tag{5.18}$$

where $\mathcal{R}_\omega$ is an operator that extracts local blocks of the Casorati matrix $\mathbf{X}$, and Ω represents the set of all blocks. To minimise the functional in Eq. (5.18), we propose a fast proximal gradient method outlined in algorithm 7.1. The blockwise singular value soft thresholding operator $\text{BSVT}^{\mathcal{S}}$ consists in applying the solution of the proximal operator of the nuclear norm individually to each block in the set Ω using the extracting operator $\mathcal{R}_\omega$.

Algorithm 5.5 Local low-rank matrix recovery via fast proximal gradient (LLR-FPG)

Input: $\mathbf{y}$, $\alpha \geq 0$, $\rho > 0$, $\Omega > 0$
Initialise: $k = 0$, $\mathbf{X}^0 = \mathbf{W}^0 = \mathbf{0}, t^0 = 1$
while stopping criterion is not met **do**
  $\mathbf{X}^{k+1} \leftarrow \text{BSVT}^{\mathcal{S}_{\rho\alpha}}_{\omega \in \Omega}(\mathbf{W}^k - \rho \mathcal{E}^*(\mathcal{E}(\mathbf{W}^k) - \mathbf{y}))$
  $t^{k+1} \leftarrow \frac{1}{2}(1 + \sqrt{1 + 4(t^k)^2})$
  $\mathbf{W}^{k+1} \leftarrow \mathbf{X}^{k+1} + \frac{t^k - 1}{t^{k+1}}(\mathbf{X}^{k+1} - \mathbf{X}^k)$
end while
Output: $\widehat{\mathbf{X}} = \mathbf{X}^k$

Reconstructions of the SL phantom (figure 5.4) from noisy, highly undersampled $(\mathbf{k}, t)$-space samples (acceleration factor of 10) using radial equispaced angle trajectory were conducted using (global) low-rank regularisation (LR-FPG, algorithm 5.2) and locally low-rank regularisation with $|\Omega| = \{4, 16\}$ (LLR-FPG, algorithm 5.5).

Figure 5.19 shows both magnitude images and NMSE errors with a distinct advantage for low-rank promoted locally. Note that we can perceive the blocks that were soft thresholded in the image for $\Omega = 16$ (bottom right). Observe the superior reconstruction results for blocks that are motionless and present only noise with no

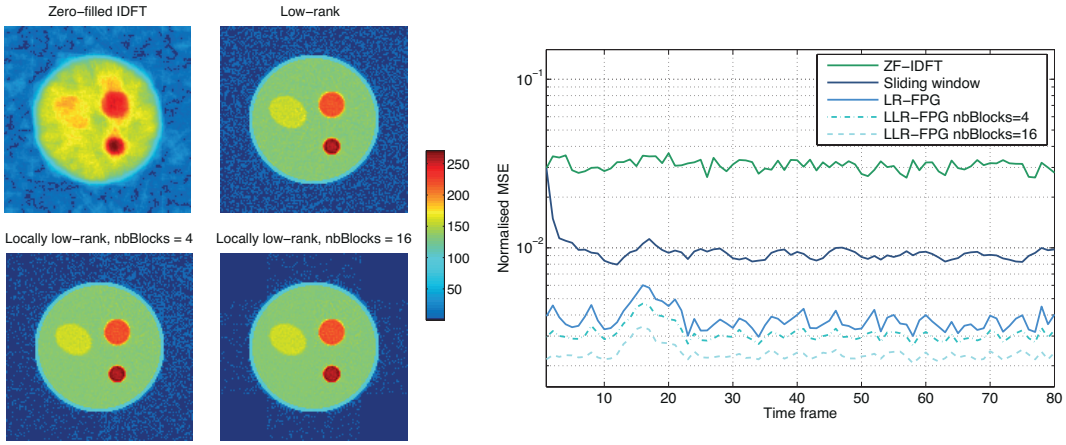

Figure 5.19: Extracted magnitude images from the reconstructed sequences (left), and NMSE (right). Note the specific colour mapping to highlight the effect of blockwise SVT (pixels with zero value have a clearly distinct colour).

signal in the original data, i.e. blocks at the corner of the image sequence. This effect is due to the fact that in this experiment we have chosen non-overlapping blocks for simplicity, but this is likely to disappear when using overlapping blocks with averaging. In this method, the number of blocks is an important question that impacts the reconstruction performance, and the best choice is yet to be determined. Computational times for LLR-FPG were as competitive as LR-FPG due to the fast optimisation scheme in algorithm 5.5 and because the blockwise SVT operator calls SVD routines on smaller matrices (although many more times than in globally low-rank). While this example clearly illustrates the advantage of encouraging low-rank locally, it should be kept in mind that higher reconstruction results may not be as substantial for dynamic sequences presenting important motion such as free-breathing data.

5.4 Discussion

5.4.1 Sparsity and low-rank prior

The Fourier transform along the temporal dimension as the sparsifying transform was chosen in this study. This transform is particularly adequate because most of the time dynamic MR signals exhibit periodicity in time. This transform is also simple to implement and very attractive computationally due to the FFT algorithm. This explains why this transform has also been used many times in other reconstruction methods.

Further work is needed to determine which sparsifying transform sparsify the most dynamic MR signals in the case of breath-hold and free-breathing data. Although various other sparsifying transforms have been used such as wavelet in time [72] or spatio-temporal total variation as in **k**-t SLR, a consistent compari-

son do not exist to our knowledge. All these transforms can be classified as fixed bases approaches because their bases are not variable, but recently, there has been interest in developing adaptive sparsifying transforms i.e., learnt from the data itself [143, 144]. The study of these adaptive sparsifying transforms in combination with low-rank prior could be of interest.

Another aspect is to think about how the low-rank prior can be exploited differently to enable better reconstruction. As we have shown in section 5.3.6, the idea of Trzasko and Manduca [142] about promoting locally low-rank instead of globally is particularly powerful. Further work in this direction should be addressed, first to combine this approach with sparsity for example using the LRS-FCS algorithm, and second to similarly find out other low-rank structures that could exhibit stronger low-rank assumption than the standard Casorati matrix.

5.4.2 Flexibility and computational times

Throughout this study, proximal gradient methods have revealed to be a flexible algorithmic framework, since they can solve efficiently sparse, low-rank and joint sparse/low-rank problems. They exploit the fact that the objective function can be split into a differentiable data fidelity term whose gradient is known and regularisation penalties whose proximal operators can be computed easily.

In this work, we have used *fast* proximal gradient methods which were based on acceleration schemes proposed by Beck and Teboulle [110] and Nesterov [109]. Thus, the developed algorithms enjoy attractive rates of convergence on the objective function, and although the actual convergence of sequence produced by these schemes is not known [100], these algorithms have been shown to converge empirically. Since these algorithms converge rapidly, they have highly competitive execution times as it has been shown in figures 5.11 and 5.12. The computational advantage would be even more attractive for imaging in higher dimensions, such as four-dimensional imaging where a volume is imaged in time.

5.4.3 Alternatives to convex optimisation

Nonconvex approaches are often motivated by the fact that they can theoretically provide better reconstruction results in contrast to convex relaxation of nonconvex problems. However, we did not make such observations in this study. Additionally, we have found nonconvex methods more difficult to handle because it is not possible to tell whether the method is performing at its best or not, since it can be stuck in local minima. This is in contrast to convex optimisation that is guaranteed to approach a global solution when the functional is minimised.

The basic greedy approach based on an iterative hard thresholding did not particularly show higher reconstruction results, although further investigation would be

needed to fully address a proper comparison.

5.4.4 Regularisation parameters

In this study, ground truths were accessible and the reconstructions were based on the minimum error according to Eq. (B.2), i.e. from a range of different regularisation parameters we have selected the reconstruction that returned the best quality. The influence of the regularisation parameters on the reconstructions is shown in figure 5.20 for the different methods. This figure gives some insights about the performance of the methods regarding the choice of regularisation parameters. In the case where multiple priors are used, these figures also show which prior help the most the reconstruction. Note that a finer selection of these parameters can potentially give different reconstruction results. In this sense, all these methods suffer from sensitivity of regularisation parameters. Automatic selection of the best regularisation parameters is still an open problem, although there has been recent work trying to tackle this for specific problems using the Stein's unbiased risk estimate, see Refs. [141, 145].

5.5 Related works

For the historical aspects, we recall that the use of accelerated gradient techniques as a faster way to solve linear discrete inverse problems with ℓ_1 minimisation through iterative soft thresholding was proposed by Beck and Teboulle with FISTA [110] in 2009. In fact, Nesterov had previously proposed a somehow more general method for minimising composite convex functions in an unpublished manuscript [109] that was proven to converge in function values with the same rate of convergence as FISTA. Following developments in low-rank matrix recovery, Toh and Yun [137] proposed in 2010 to adapt the FISTA scheme for nuclear norm minimisation. Combettes and Pesquet made a significant contribution by proposing the proximal splitting methods [100], as a generalised framework to solve similar type of convex optimisation problems.

To our knowledge, the use of fast proximal gradient methods for low-rank and/or sparsity reconstruction in dynamic imaging has received limited attention in the MR community. Many proposed MRI reconstruction algorithms for low-rank and sparsity rely on other optimisation techniques which have been described in section 3.4.

5.6 Conclusion

In this chapter, we have primarily provided simple, efficient and fast algorithms to solve low-rank and sparse MRI reconstruction problems via convex optimisation

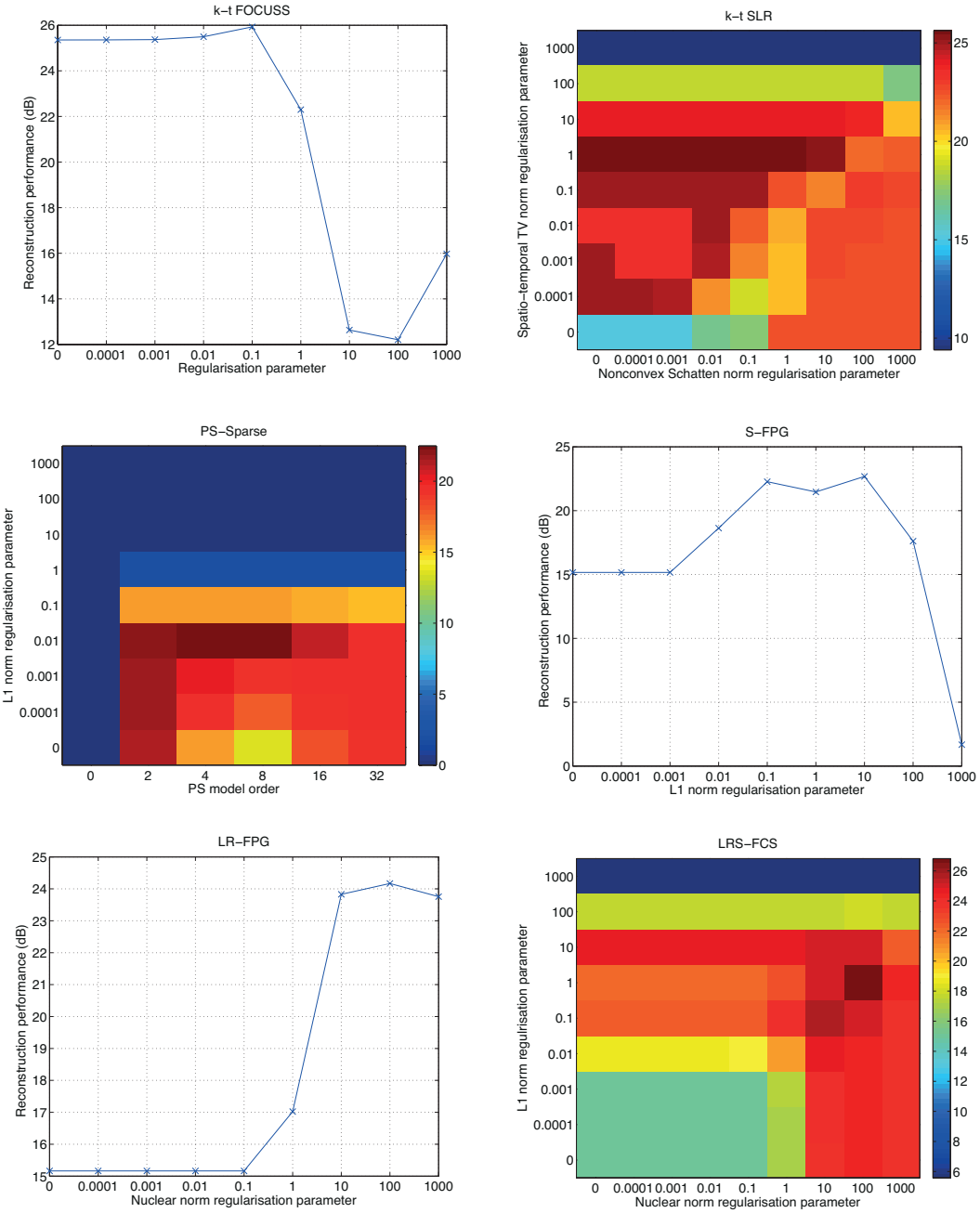

Figure 5.20: Influence of the regularisation parameters for the different methods, showing here the SL phantom and golden angle radial sampling.

and proximal gradient methods. We also have provided numerous comparisons in terms of datasets, sampling patterns and state of the art reconstruction methods. We have confirmed previous studies showing that exploiting simultaneously low-rank and sparsity could benefit the reconstruction of dynamic MR images from partial data compared to individually sparse or low-rank, but this study has also revealed that low-rank matrix recovery on its own could be a serious alternative to sparse minimisation. Additionally, the developed algorithms were all training-free, i.e. no training set or low-resolution estimate (as in **k**-t FOCUSS or PS-sparse) were required to reconstruct data.

One of the developed convex algorithm that exploits both low-rank structure and sparseness in the the temporal Fourier domain (LRS-FCS) has proved to consistently generate higher quality results in much more competitive running times than state of art low-rank and sparse reconstruction methods. Having both strong reconstruction accuracy and low computational costs, this method reveals to be a highly competitive reconstruction algorithm against state of the art methods.

Chapter 6

Joint reconstruction–separation via matrix decomposition

Contents

6.1 Introduction

In this chapter, we introduce a joint reconstruction–separation method from partial $(\mathbf{k}, t)$-space measurements that reconstructs and inherently separates into two components the information in the dynamic scene.

The reconstruction is based on the low-rank plus sparse (L+S) matrix decomposition model, also known as robust principal components analysis (RPCA). The motivations for using this model are twofold. First, RPCA applied to a set of dynamic images results in two distinct low-rank and sparse components that capture respectively the background and dynamic information [56]. Second, it has been shown in multiple studies [56,58–62] that it is possible to recover both low-rank and sparse components from only a fraction of observations under some specific assumptions. In our proposed approach, we assume that the Casorati matrix formulation of the dynamic MRI sequence can be expressed as a linear combination of a low-rank plus sparse component, and at the same time that this prior information is strong enough to be able to reconstruct images from partial $(\mathbf{k}, t)$-space samples.

The contribution of this work is to propose a joint reconstruction–separation method from sub-Nyquist $(\mathbf{k}, t)$-space samples that intrinsically reconstructs and separates data through the L+S matrix decomposition model. In some sense, this approach goes beyond traditional reconstruction techniques for accelerated MRI because it is a first step towards methods that could anticipate or help in the interpretation of images directly from the reconstruction procedure. The method is named $\mathbf{k}$-t RPCA, since it involves the reconstruction of images from partial acquired $(\mathbf{k}, t)$-space samples using the RPCA model. Work presented in this chapter has been previously published in Refs. [90, 146].

This chapter is organised as follows. The proposed method is presented in section 6.2 where we also briefly review the robust principal component analysis method. Numerical simulations are shown in section 6.3. Sections 6.4, 6.5 and 6.6 respectively present the discussion, related works and the conclusion of this study.

6.2 Method

6.2.1 Robust principal component analysis

Robust principal component analysis (RPCA) [55–57], also referred to as *low-rank plus sparse (matrix) decomposition* (L+S), was initially motivated by the fact that the matrix representation of some system in many applications can be characterised as being composed of a low-rank and a sparse matrix.

More formally, the goal of RPCA is to decompose a given matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ into its low-rank $\mathbf{L}$ and sparse $\mathbf{S}$ components such that $\mathbf{X} = \mathbf{L} + \mathbf{S}$, as illustrated in figure 6.1. In other words, one wants to find the lowest rank of $\mathbf{L}$ that could have generated the observed data, subject to the constraint that $\mathbf{S}$ is sparse and naturally that $\mathbf{L} + \mathbf{S} = \mathbf{X}$. This can be expressed as a minimisation problem,

$$\min_{\mathbf{L},\mathbf{S}} \operatorname{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0 \quad \text{s.t.} \quad \mathbf{L} + \mathbf{S} = \mathbf{X}. \tag{6.1}$$

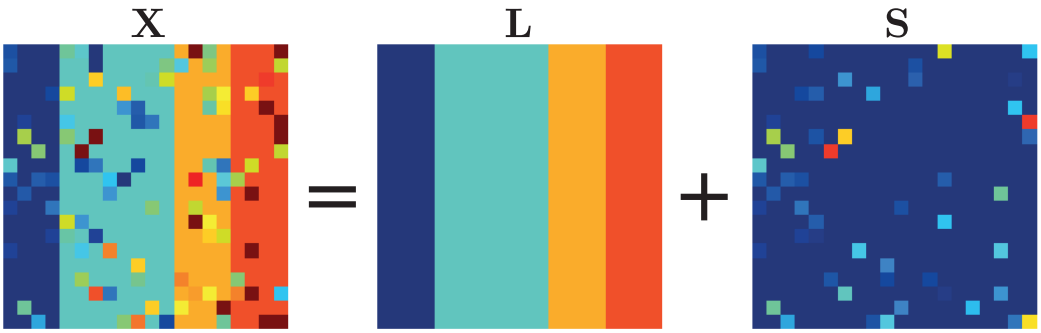

Figure 6.1: Schematic RPCA decomposition. Given a matrix $\mathbf{X}$ that is neither low-rank nor sparse, RPCA estimates low-rank $\mathbf{L}$ and sparse $\mathbf{S}$ matrices such that $\mathbf{X} = \mathbf{L}+\mathbf{S}$.

As it has been explained in chapter 3, problem (6.1) is intractable due to the rank operator and ℓ_0 pseudonorm. The convex relaxation technique can be used by employing instead respectively the nuclear and ℓ_1 norms,

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathbf{L}+\mathbf{S} = \mathbf{X}. \tag{6.2}$$

Solving the above problem can be interpreted as an attempt to make a more robust version of the standard principal component analysis (PCA) regarding grossly corrupted observations (outliers), since the sparse component will tend to capture these outliers. Note that the low-rank plus sparse decomposition is not unique if the given matrix is both low-rank and sparse, since both components would be seen interchangeably as either low-rank or sparse in this case (for example, a matrix that has only one nonzero element).

RPCA is solved efficiently via the alternating direction methods of multipliers (ADMM) (as described in section 4.3) because of the intrinsic separable structure that appears in both the objective function and the constraint. First, the augmented Lagrangian function of (6.2) must be expressed,

$$\mathcal{L}_\delta^A(\mathbf{L},\mathbf{S},\mathbf{Z}) = \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 + \langle \mathbf{Z}, \mathbf{L}+\mathbf{S}-\mathbf{X}\rangle + \frac{\delta}{2}\|\mathbf{L}+\mathbf{S}-\mathbf{X}\|_F^2, \tag{6.3}$$

where $\langle .,. \rangle$ denotes the trace inner product, $\mathbf{Z}$ is the Lagrange multiplier of the linear constraint and δ is the penalty parameter. ADMM can be formulated as

$$\begin{aligned}
\mathbf{L}^{k+1} &\leftarrow \arg\min_{\mathbf{L}} \mathcal{L}_\delta^A(\mathbf{L},\mathbf{S}^k,\mathbf{Z}^k) \\
\mathbf{S}^{k+1} &\leftarrow \arg\min_{\mathbf{S}} \mathcal{L}_\delta^A(\mathbf{L}^{k+1},\mathbf{S},\mathbf{Z}^k) \\
\mathbf{Z}^{k+1} &\leftarrow \mathbf{Z}^k + \delta(\mathbf{L}^{k+1}+\mathbf{S}^{k+1}-\mathbf{X}).
\end{aligned} \tag{6.4}$$

The scaled form of ADMM can also be derived,

$$\mathbf{L}^{k+1} \leftarrow \arg\min_{\mathbf{L}} \frac{1}{\delta}\|\mathbf{L}\|_* + \frac{1}{2}\|\mathbf{L} - [\mathbf{X} - \mathbf{S}^k - \mathbf{U}^k]\|_F^2 \tag{6.5a}$$

$$\mathbf{S}^{k+1} \leftarrow \arg\min_{\mathbf{S}} \frac{\lambda}{\delta}\|\mathbf{S}\|_1 + \frac{1}{2}\|\mathbf{S} - [\mathbf{X} - \mathbf{L}^{k+1} - \mathbf{U}^k]\|_F^2 \tag{6.5b}$$

$$\mathbf{U}^{k+1} \leftarrow \mathbf{U}^k + \mathbf{L}^{k+1} + \mathbf{S}^{k+1} - \mathbf{X}, \tag{6.5c}$$

where $\mathbf{U}^k = (1/\delta)\mathbf{Z}^k$. Since (6.5a) and (6.5b) are proximal operators of respectively the nuclear norm and ℓ_1 norm evaluated at $\mathbf{X} - \mathbf{S}^k - \mathbf{U}^k$ and $\mathbf{X} - \mathbf{L}^{k+1} - \mathbf{U}^k$, a proximal version can be established that is described in algorithm 6.1. The penalty parameter δ can be fixed to the specific value $\delta = (MN)/(4\|\mathbf{X}\|_1)$ as suggested in Refs. [56, 57] although another strategy is to update it dynamically [53].

Algorithm 6.1 Robust principal component analysis (RPCA)

Input: $\mathbf{X}$, $\lambda > 0$
Initialize: $k = 0$, $\mathbf{S}^0 = \mathbf{U}^0 = \mathbf{0}$, $\delta = \frac{1}{4}\frac{MN}{\|\mathbf{X}\|_1}$
while stopping criterion is not met **do**
  $\mathbf{L}^{k+1} \leftarrow \text{prox}_{(1/\delta)\|.\|_*}(\mathbf{X} - \mathbf{S}^k - \mathbf{U}^k)$
  $\mathbf{S}^{k+1} \leftarrow \text{prox}_{(\lambda/\delta)\|.\|_1}(\mathbf{X} - \mathbf{L}^{k+1} - \mathbf{U}^k)$
  $\mathbf{U}^{k+1} \leftarrow \mathbf{U}^k + \mathbf{L}^{k+1} + \mathbf{S}^{k+1} - \mathbf{X}$
end while
Output: $\widehat{\mathbf{L}} = \mathbf{L}^k$, $\widehat{\mathbf{S}} = \mathbf{S}^k$

The parameter λ in (6.2) plays the important role of a trade-off between how much the low-rank component gets "low-rank" and how much the sparse component gets "sparse". Candès et al. [56] suggested the theoretically supported value

$$\lambda = \max(M, N)^{-1/2}. \tag{6.6}$$

Note this value depends on the dimensions of the original matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$. In general, this choice offers a reasonable separation in between low-rank and sparse components, although multiplying this value by a scaling factor can help to tailor the separation to a given application. For that reason, we define the following expression that we will use later,

$$\lambda_\rho = \rho \max(M, N)^{-1/2} \tag{6.7}$$

where $\rho > 0$ is the scaling factor.

To apply RPCA on a dynamic MRI sequence of dimensions $N_x \times N_y$ with N_t frames, the Casorati matrix whose columns represent vectorised MR images of the sequence must be formed, see Eq. (3.59). An example of RPCA applied to a breath-hold cardiac MRI sequence is shown in figure 6.2 for a specific value of λ_ρ. Physiologically, the low-rank part appears as a static component (background) while the sparse component captures dynamics (motion), in this particular case mostly

heartbeats.

6.2.2 Joint reconstruction–separation

In this work, we are interested in the the low-rank plus sparse decomposition not only from the separation aspect, but also from a reconstruction perspective. Thus, we formulate RPCA for undersampled dynamic MRI as the following convex problem,

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda_\rho \|\mathcal{F}_t(\mathbf{S})\|_1 \quad \text{s.t.} \quad \mathcal{E}(\mathbf{L}+\mathbf{S}) = \mathbf{y}, \tag{6.8}$$

where $\mathcal{E} : \mathbb{C}^{N_x N_y \times N_t} \to \mathbb{C}^M$ represents the MRI encoding operator modelling both the random sub-Nyquist sampling and Fourier transform, $\mathcal{F}_t$ denotes the Fourier transform operator along the temporal dimension as in Eq. (3.65), λ_ρ is the decomposition parameter[1] as defined in Eq. (6.7), and $\mathbf{y} \in \mathbb{C}^M$ represents the $(\mathbf{k}, t)$-space undersampled data.

The above formulation assumes that the dynamic imaging data have the property of being separable into an approximately low-rank and approximately sparse components. The additional operator $\mathcal{F}_t$ applied to the sparse component can be justified by the fact that the proposed method deals with the reconstruction of undersampled data: the temporal Fourier transform is known to improve sparsity in many dynamic reconstruction methods (e.g. Refs. [72, 73]). This is beneficial since it is assumed in such studies that the more the signal is sparse, the higher the undersampling ratio can be. An illustration is provided in figure 6.3.

Note that the associated AL function of problem (6.8) is

$$\mathcal{L}_\delta^A(\mathbf{L},\mathbf{S},\mathbf{Z}) = \|\mathbf{L}\|_* + \lambda_\rho \|\mathcal{F}_t(\mathbf{S})\|_1 + \Re\{\langle \mathbf{Z}, \mathcal{E}(\mathbf{L}+\mathbf{S}) - \mathbf{y}\rangle\} + \frac{\delta}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S}) - \mathbf{y}\|_2^2, \tag{6.9}$$

and the ADMM takes the following formulation,

$$\mathbf{L}^{k+1} \leftarrow \arg\min_{\mathbf{L}} \mathcal{L}_\delta^A(\mathbf{L}, \mathbf{S}^k, \mathbf{Z}^k) \tag{6.10a}$$

$$\mathbf{S}^{k+1} \leftarrow \arg\min_{\mathbf{S}} \mathcal{L}_\delta^A(\mathbf{L}^{k+1}, \mathbf{S}, \mathbf{Z}^k) \tag{6.10b}$$

$$\mathbf{Z}^{k+1} \leftarrow \mathbf{Z}^k + \delta(\mathcal{E}(\mathbf{L}^{k+1} + \mathbf{S}^{k+1}) - \mathbf{y}). \tag{6.10c}$$

However, alternating steps (6.10a) and (6.10b) cannot be solved easily in particular due to the presence of the MRI operator $\mathcal{E}$. Indeed, developing step (6.10a) for

[1] The decomposition parameter can be seen interchangeably as either λ_ρ or ρ. In the following, we will mainly refer to it as ρ because it is easier to interpret since it can be seen as a scaling parameter that does not depend on the matrix dimensions.

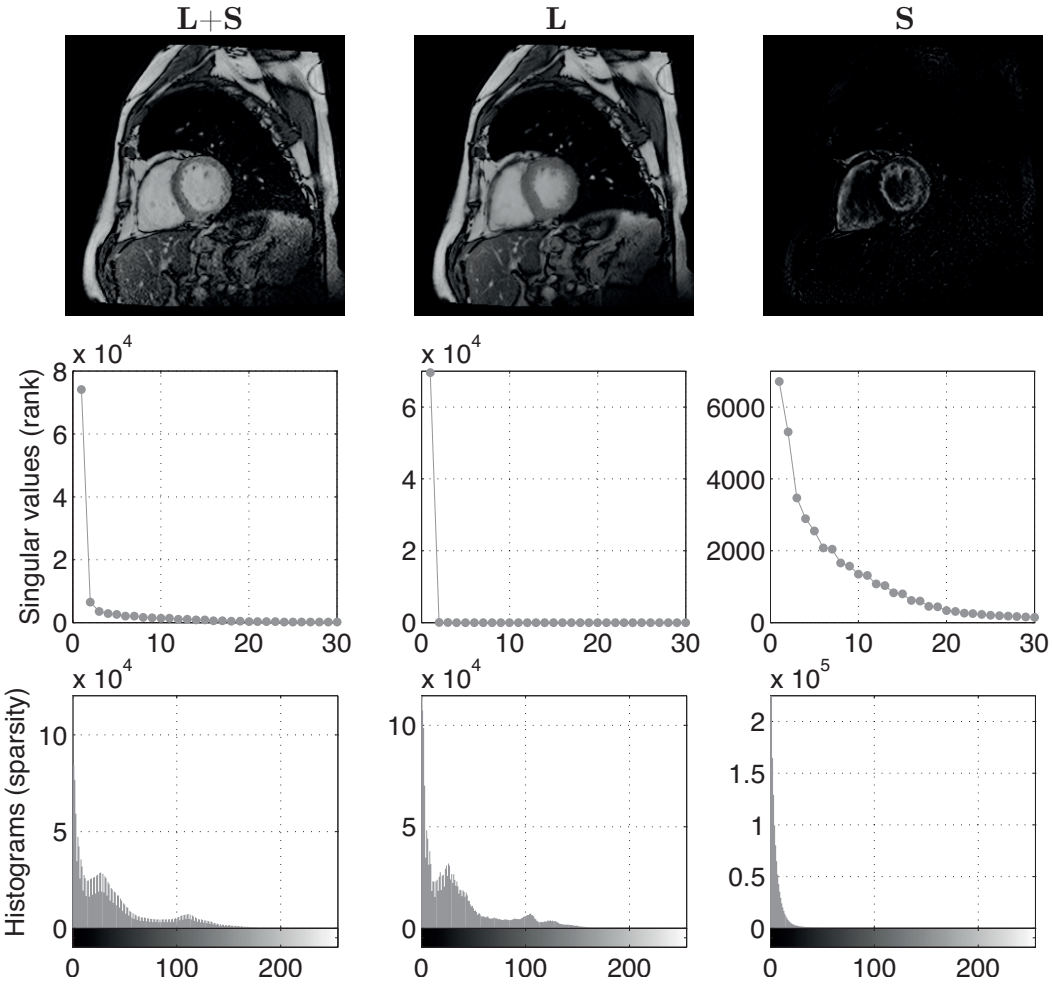

Figure 6.2: RPCA on a breath-hold cardiac cine MRI sequence with $N_t = 30$ (showing only a single frames from the sequence on the top figure). Algorithm 6.1 with $\rho = 0.5$ in Eq. (6.7) was used to generate figures in this example. The decomposition resulted in a rank-1 matrix for the low-rank part as shown by the only nonzero singular value, while the sparse component does not have a low rank because most of its singular values are not close to zero. It can be seen on the corresponding images and histograms that the sparse component is much more sparse than the low-rank one.

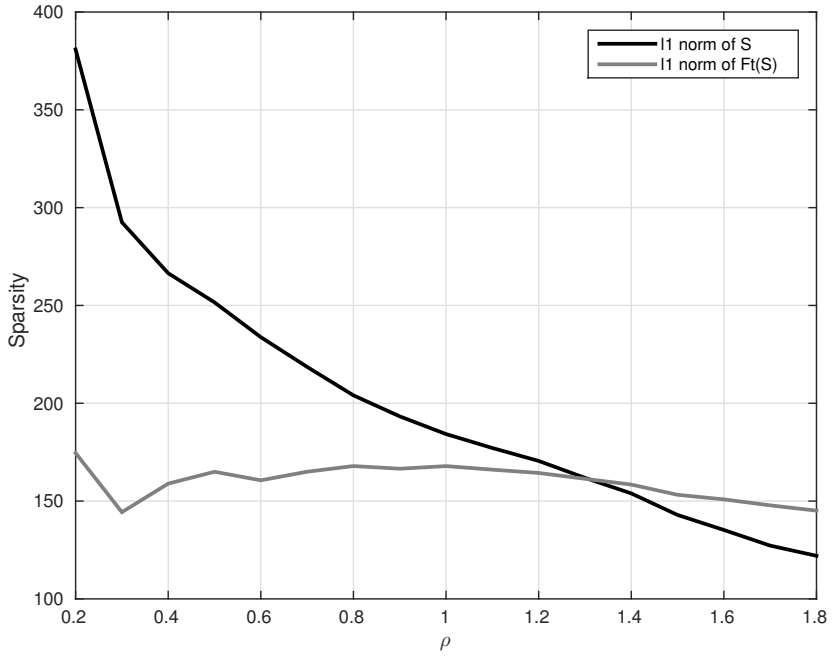

Figure 6.3: Effect of the additional temporal Fourier transform on the sparse component using different decomposition parameter ρ. Gray curve shows the normalised ℓ_1 norm of temporal Fourier transform of the sparse component ($\|\mathcal{F}_t(\mathbf{S})\|_1/\|\mathbf{S}\|_F$), black curve shows the normalised ℓ_1 norm of the sparse component ($\|\mathbf{S}\|_1/\|\mathbf{S}\|_F$). The additional temporal Fourier operator can generally help sparsifying the signal when $\mathbf{S}$ is not particularly sparse, e.g. $\rho \in (0, 1.2]$. This figure has been generated using algorithm 6.1 and the numerical phantom with a combination of motion and intensity changes (section 6.3.2).

example leads to

$$\begin{aligned}\mathbf{L}^{k+1} &\leftarrow \arg\min_{\mathbf{L}} \mathcal{L}_{\delta}^{A}(\mathbf{L}, \mathbf{S}^k, \mathbf{Z}^k) \\ &\leftarrow \arg\min_{\mathbf{L}} \Big\{ \|\mathbf{L}\|_* + \Re\{\mathbf{Z}^{k\mathrm{H}}(\mathcal{E}(\mathbf{L}+\mathbf{S}^k)-\mathbf{y})\} + \frac{\delta}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S}^k)-\mathbf{y}\|_2^2 \Big\} \\ &\leftarrow \arg\min_{\mathbf{L}} \Big\{ \|\mathbf{L}\|_* + \Re\{\mathbf{Z}^{k\mathrm{H}}\mathcal{E}(\mathbf{L}+\mathbf{S}^k)\} - \Re\{\mathbf{Z}^{k\mathrm{H}}\mathbf{y}\} + \frac{\delta}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S}^k)-\mathbf{y}\|_2^2 \Big\} \\ &\leftarrow \arg\min_{\mathbf{L}} \Big\{ \|\mathbf{L}\|_* + \Re\{\mathbf{Z}^{k\mathrm{H}}\mathcal{E}(\mathbf{L}+\mathbf{S}^k)\} + \frac{\delta}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S}^k)-\mathbf{y}\|_2^2 \Big\},\end{aligned} \tag{6.11}$$

which is a minimisation problem difficult to solve. To overcome this, image reconstruction is first reformulated as an unconstrained minimisation of the convex objective function $F(\mathbf{L},\mathbf{S}) : \mathbb{C}^{N_x N_y \times N_t} \times \mathbb{C}^{N_x N_y \times N_t} \to \mathbb{R}$,

$$\{\mathbf{L}^\star, \mathbf{S}^\star\} = \arg\min_{\mathbf{L},\mathbf{S}} \Big\{ F(\mathbf{L},\mathbf{S}) \equiv \frac{1}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S})-\mathbf{y}\|_2^2 + \mu\Big(\|\mathbf{L}\|_* + \lambda_\rho \|\mathcal{F}_t(\mathbf{S})\|_1 \Big) \Big\}, \tag{6.12}$$

where μ represents the regularisation parameter.

6.2.3 Image reconstruction algorithm

To minimise $F(\mathbf{L},\mathbf{S})$ as defined in (6.12), an algorithm is derived based on ADMM. First, the variable splitting procedure is used,

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{L},\mathbf{S}} \frac{1}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S})-\mathbf{y}\|_2^2 + \mu(\|\mathbf{P}\|_* + \lambda_\rho\|\mathbf{Q}\|_1) \quad \text{s.t.} \quad \mathbf{P}=\mathbf{L}, \mathbf{Q}=\mathcal{F}_t(\mathbf{S}). \tag{6.13}$$

The associated AL function reads

$$\begin{aligned}\mathcal{L}_{\delta_1,\delta_2}^{A}(\mathbf{P},\mathbf{Q},\mathbf{L},\mathbf{S}) = \frac{1}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S})-\mathbf{y}\|_2^2 + \mu\|\mathbf{P}\|_* + \mu\lambda_\rho\|\mathbf{Q}\|_1 \\ + \Re\{\langle \mathbf{Z}_1, \mathbf{L}-\mathbf{P}\rangle\} + \frac{\delta_1}{2}\|\mathbf{L}-\mathbf{P}\|_2^2 \\ + \Re\{\langle \mathbf{Z}_2, \mathcal{F}_t(\mathbf{S})-\mathbf{Q}\rangle\} + \frac{\delta_2}{2}\|\mathcal{F}_t(\mathbf{S})-\mathbf{Q}\|_2^2,\end{aligned} \tag{6.14}$$

where $\mathbf{Z}_i$ are Lagrangian multipliers. Combining the linear and quadratic terms in the augmented Lagrangian function and ignoring constants irrelevant to optimisation, Eq. (6.14) can also be written as

$$\begin{aligned}\mathcal{L}_{\delta_1,\delta_2}^{A}(\mathbf{P},\mathbf{Q},\mathbf{L},\mathbf{S}) = \frac{1}{2}\|\mathcal{E}(\mathbf{L}+\mathbf{S})-\mathbf{y}\|_2^2 + \mu\|\mathbf{P}\|_* + \mu\lambda_\rho\|\mathbf{Q}\|_1 \\ + \frac{\delta_1}{2}\|\mathbf{L}+\delta_1^{-1}\mathbf{Z}_1-\mathbf{P}\|_2^2 + \frac{\delta_2}{2}\|\mathcal{F}_t(\mathbf{S})+\delta_2^{-1}\mathbf{Z}_2-\mathbf{Q}\|_2^2.\end{aligned} \tag{6.15}$$

ADMM minimises $\mathcal{L}^A_{\delta_1,\delta_2}$ over $\mathbf{P}, \mathbf{Q}, \mathbf{L}$ and $\mathbf{S}$ separately, which leads to this iterative procedure,

$$\begin{aligned}
\mathbf{P}^{k+1} &\leftarrow \arg\min_{\mathbf{P}} \mathcal{L}^A_{\delta_1,\delta_2}(\mathbf{P}, \mathbf{Q}^k, \mathbf{L}^k, \mathbf{S}^k) \\
\mathbf{Q}^{k+1} &\leftarrow \arg\min_{\mathbf{Q}} \mathcal{L}^A_{\delta_1,\delta_2}(\mathbf{P}^{k+1}, \mathbf{Q}, \mathbf{L}^k, \mathbf{S}^k) \\
\mathbf{L}^{k+1} &\leftarrow \arg\min_{\mathbf{L}} \mathcal{L}^A_{\delta_1,\delta_2}(\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}, \mathbf{L}, \mathbf{S}^k) \\
\mathbf{S}^{k+1} &\leftarrow \arg\min_{\mathbf{S}} \mathcal{L}^A_{\delta_1,\delta_2}(\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}, \mathbf{L}^{k+1}, \mathbf{S}) \\
\mathbf{Z}_1^{k+1} &\leftarrow \mathbf{Z}_1^k + \delta_1(\mathbf{L}^{k+1} - \mathbf{P}^{k+1}) \\
\mathbf{Z}_2^{k+1} &\leftarrow \mathbf{Z}_2^k + \delta_2(\mathcal{F}_t(\mathbf{S}^{k+1}) - \mathbf{Q}^{k+1}).
\end{aligned} \tag{6.16}$$

These sub-problems result either in proximal operators that we have encountered in chapter 4, or quadratic problems leading to linear system of equations,

$$\mathbf{P}^{k+1} \leftarrow \arg\min_{\mathbf{P}} \frac{1}{2}\|\mathbf{L}^k + \delta_1^{-1}\mathbf{Z}_1^k - \mathbf{P}\|_2^2 + \frac{\mu}{\delta_1}\|\mathbf{P}\|_* \tag{6.17a}$$

$$\leftarrow \text{prox}_{\mu/\delta_1\|.\|_*}(\mathbf{L}^k + \delta_1^{-1}\mathbf{Z}_1^k) \tag{6.17b}$$

$$\leftarrow \text{SVT}^{\mathcal{S}_{\mu/\delta_1}}(\mathbf{L}^k + \delta_1^{-1}\mathbf{Z}_1^k) \tag{6.17c}$$

$$\mathbf{Q}^{k+1} \leftarrow \arg\min_{\mathbf{Q}} \frac{1}{2}\|\mathcal{F}_t(\mathbf{S}^k) + \delta_2^{-1}\mathbf{Z}_2^k - \mathbf{Q}\|_2^2 + \frac{\mu\lambda_\rho}{\delta_2}\|\mathbf{Q}\|_1 \tag{6.17d}$$

$$\leftarrow \text{prox}_{(\mu\lambda_\rho)/\delta_2\|.\|_1}(\mathcal{F}_t(\mathbf{S}^k) + \delta_2^{-1}\mathbf{Z}_2^k) \tag{6.17e}$$

$$\leftarrow \mathcal{S}_{(\mu\lambda_\rho)/\delta_2}(\mathcal{F}_t(\mathbf{S}^k) + \delta_2^{-1}\mathbf{Z}_2^k) \tag{6.17f}$$

$$\mathbf{L}^{k+1} \leftarrow \arg\min_{\mathbf{L}} \frac{1}{2}\|\mathcal{E}(\mathbf{L} + \mathbf{S}^k) - \mathbf{y}\|_2^2 + \frac{\delta_1}{2}\|\mathbf{L} + \delta_1^{-1}\mathbf{Z}_1^k - \mathbf{P}^{k+1}\|_2^2 \tag{6.17g}$$

$$\leftarrow (\mathcal{E}^*\mathcal{E} + \delta_1\mathbf{I})^{-1}(\mathcal{E}^*\mathbf{y} + \delta_1\mathbf{P}^{k+1} - \mathbf{Z}_1^k - \mathcal{E}^*\mathcal{E}\mathbf{S}^k) \tag{6.17h}$$

$$\mathbf{S}^{k+1} \leftarrow \arg\min_{\mathbf{S}} \frac{1}{2}\|\mathcal{E}(\mathbf{L}^{k+1} + \mathbf{S}) - \mathbf{y}\|_2^2 + \frac{\delta_2}{2}\|\mathcal{F}_t(\mathbf{S}) + \delta_2^{-1}\mathbf{Z}_2^k - \mathbf{Q}^{k+1}\|_2^2 \tag{6.17i}$$

$$\leftarrow (\mathcal{E}^*\mathcal{E} + \delta_2\mathbf{I})^{-1}(\mathcal{E}^*\mathbf{y} + \mathcal{F}_t{}^*(\delta_2\mathbf{Q}^{k+1} - \mathbf{Z}_2) - \mathcal{E}^*\mathcal{E}\mathbf{L}^{k+1}). \tag{6.17j}$$

Based on these closed-form solutions, the image reconstruction procedure is derived in algorithm 6.2. Some additional modifications include the penalty parameters δ_1 and δ_2 which are both set and fixed to 1 (although as in standard RPCA they could be updated dynamically) and the re-enforcement of the decomposition at the beginning of the iterative process: the term $\mathbf{L}^k$ in (6.17c) is replaced by $\mathbf{X}^k - \mathbf{S}^k$, and similarly the term $\mathbf{S}^k$ in (6.17f) is replaced by $\mathbf{X}^k - \mathbf{L}^k$. The latter modification is also useful to reduce the number of iterations needed to obtain a satisfying decomposition.

The algorithm is stopped if a maximum number of 200 iterations is reached or if the estimated matrix does not change significantly any more between two iterations, i.e. if

$$\frac{\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F}{\|\mathbf{X}^k\|_F} \leq \texttt{tol}, \tag{6.18}$$

where `tol` is the tolerance set to a small value such as 10^{-6}. Another possibility would be to stop the algorithm when the objective function does not decrease significantly any more, as in Eq. (5.12).

Algorithm 6.2 Dynamic MR image reconstruction–separation via low-rank plus sparse prior (**k**-t RPCA)

Input: $\mathbf{y}$, $\mu, \lambda_\rho > 0$
Initialize: $k = 0$, $\mathbf{X}^0 = \mathbf{L}^0 = \mathcal{E}^*(\mathbf{y})$, $\mathbf{S}^0 = \mathbf{Z}_i^0 = \mathbf{0}$
while stopping criterion is not met **do**
$\mathbf{P}^{k+1} \leftarrow \mathrm{SVT}^{\mathcal{S}_\mu}(\mathbf{X}^k - \mathbf{S}^k + \mathbf{Z}_1^k)$
$\mathbf{Q}^{k+1} \leftarrow \mathcal{S}_{\mu\lambda_\rho}(\mathcal{F}_t(\mathbf{X}^k - \mathbf{L}^k) + \mathbf{Z}_2^k)$
$\mathbf{L}^{k+1} \leftarrow (\mathcal{E}^*\mathcal{E} + \mathbf{I})^{-1}(\mathcal{E}^*(\mathbf{y}) + \mathbf{P}^{k+1} - \mathbf{Z}_1^k - \mathcal{E}^*(\mathcal{E}(\mathbf{S}^k)))$
$\mathbf{S}^{k+1} \leftarrow (\mathcal{E}^*\mathcal{E} + \mathbf{I})^{-1}(\mathcal{E}^*(\mathbf{y}) + \mathcal{F}_t{}^*(\mathbf{Q}^{k+1} - \mathbf{Z}_2^k) - \mathcal{E}^*(\mathcal{E}(\mathbf{L}^{k+1})))$
$\mathbf{Z}_1^{k+1} \leftarrow \mathbf{Z}_1^k + \mathbf{L}^{k+1} - \mathbf{P}^{k+1}$
$\mathbf{Z}_2^{k+1} \leftarrow \mathbf{Z}_2^k + \mathcal{F}_t(\mathbf{S}^{k+1}) - \mathbf{Q}^{k+1}$
$\mathbf{X}^{k+1} \leftarrow \mathbf{L}^{k+1} + \mathbf{S}^{k+1}$
end while
Output: $\widehat{\mathbf{L}} = \mathbf{L}^k$, $\widehat{\mathbf{S}} = \mathbf{S}^k$

6.2.4 Sampling considerations

In section 5.2.5 we have explained that sampling patterns in MRI are subject to hardware constraints. For this study, we use and compare two sampling strategies based on polynomial variable density sampling and pseudo-radial sampling. An illustration of these sampling patterns is shown in figure 6.4.

6.3 Numerical simulations

6.3.1 Framework

Experiments were run in Matlab on a Linux platform. Intensity of data were normalised between values 0 and 255 prior to any processing. Simulated data were created directly in the image domain and *in vivo* data were based on magnitude-reconstructed images from an MR scanner. Datasets were then undersampled retrospectively using a polynomial variable density or pseudo-radial sampling schemes as shown in figure 6.4. To obtain more realistic simulations, Gaussian noise was added explicitly on each real and imaginary channel of the undersampled data in all experiments with zero mean and a standard deviation $\sigma = 3$. We refer the reader to appendix B for the error metrics used to evaluate the performance of the reconstructions.

A zero-filled inverse Fourier transform and a sliding window reconstruction using a zeroth-order hold technique [69] are included, mainly to illustrate the level

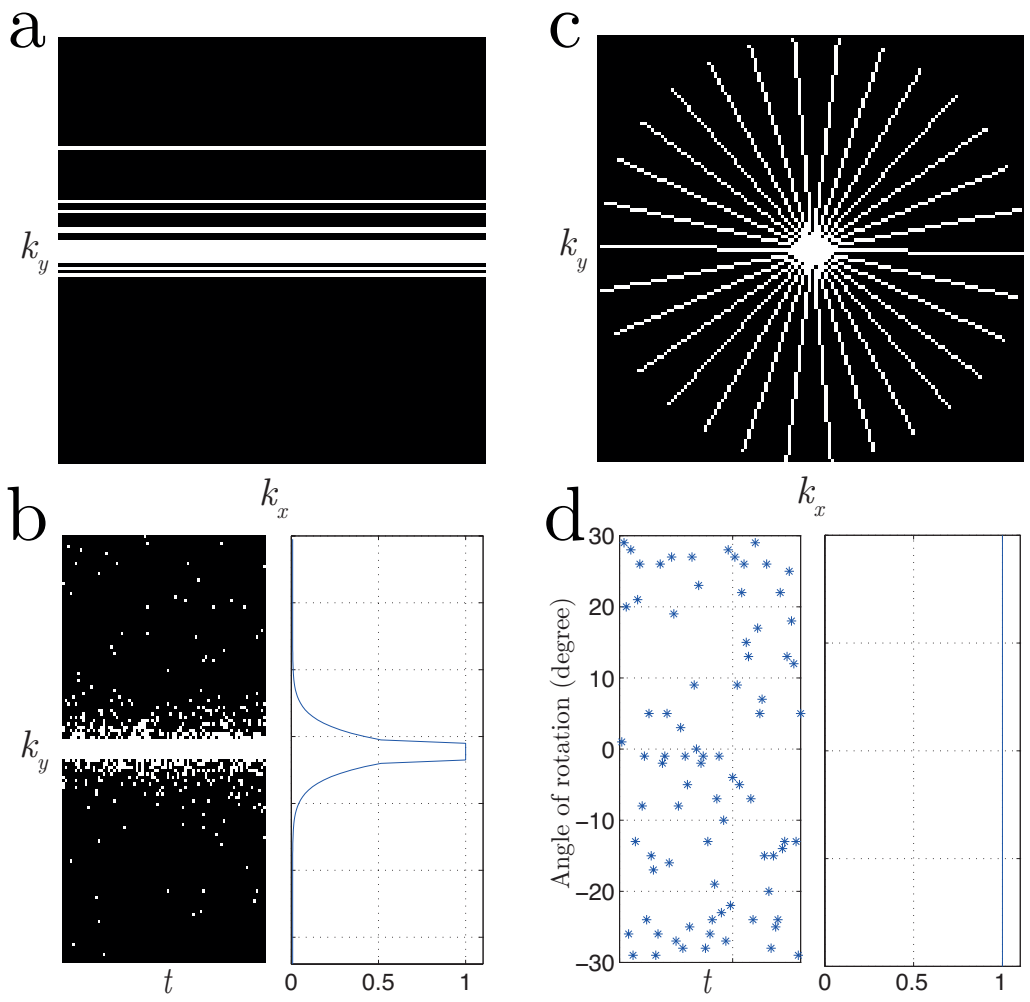

Figure 6.4: (a) One time frame acquisition pattern for polynomial variable density sampling and (b) the (k_y, t)-space sampling pattern (left) with its associated probability density function (right). (c) One time frame acquisition pattern for pseudo-radial sampling and (d) the (random) angles of rotation in time (left) with the associated uniform probability density function (right).

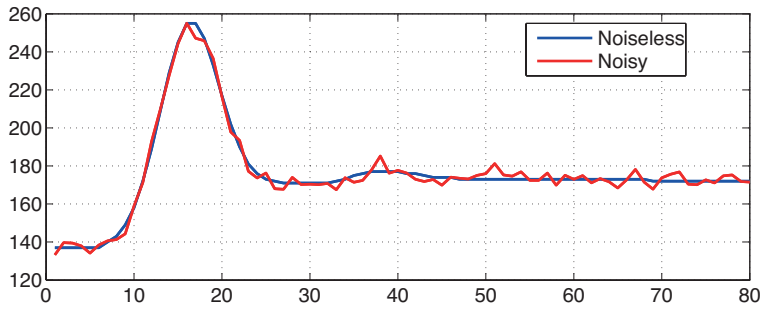

Figure 6.5: Modelling local intensity changes (showing here pixel intensity values in time) as the uptake and washout of a contrast agent using the modified Tofts model.

of undersampling. Comparisons with dynamic MR reconstruction methods **k**-t FOCUSS [73–75] and **k**-t SLR [82] are provided. These methods were described previously in section 3.4. **k**-t FOCUSS is implemented with 40 inner iterations (conjugate gradient step), 2 outer iterations (FOCUSS step) and weighting matrix power factor of 0.5 (cf. section 3.4.2). The low-resolution initial estimate is obtained by using a zero-filled inverse Fourier transform using the low-frequency samples. For **k**-t SLR, there are parameters to tune related to the continuation strategy of the optimisation algorithm that are used to improve the convergence rate. These parameters are set to suggested values provided in the **k**-t SLR package (penalty parameters $\beta_1 = \beta_2 = 10^{-7}$ for Schatten and TV norms; penalty parameters incrementation both set to 25 in the outer loop; maximum number of 50 inner and 9 outer iterations).

In **k**-t FOCUSS, one regularisation parameter can be tuned to control the stability of the solution under noisy conditions. Here, reconstructions with a different regularisation parameter selected from a range of values are computed and the best one is selected in accordance with Eq. (B.2). In **k**-t SLR, different regularisation parameters α and β for respectively the Schatten norm and spatio-temporal TV norm are tested, and the best reconstruction is selected according to Eq. (B.2). A similar strategy for **k**-t RPCA is employed by varying both the regularisation μ and decomposition ρ parameters. Note that ρ can be fixed to obtain a specific type of decomposition, although it should be noted that it may also affect the reconstruction results, which is discussed in section 6.4.

6.3.2 Reconstruction results

This section presents reconstruction results where the separation is not of particular interest, but proves to be a strong enough a priori information to remain competitive against state of the art methods. In these experiments, ρ in **k**-t RPCA is automatically selected to return the best reconstruction.

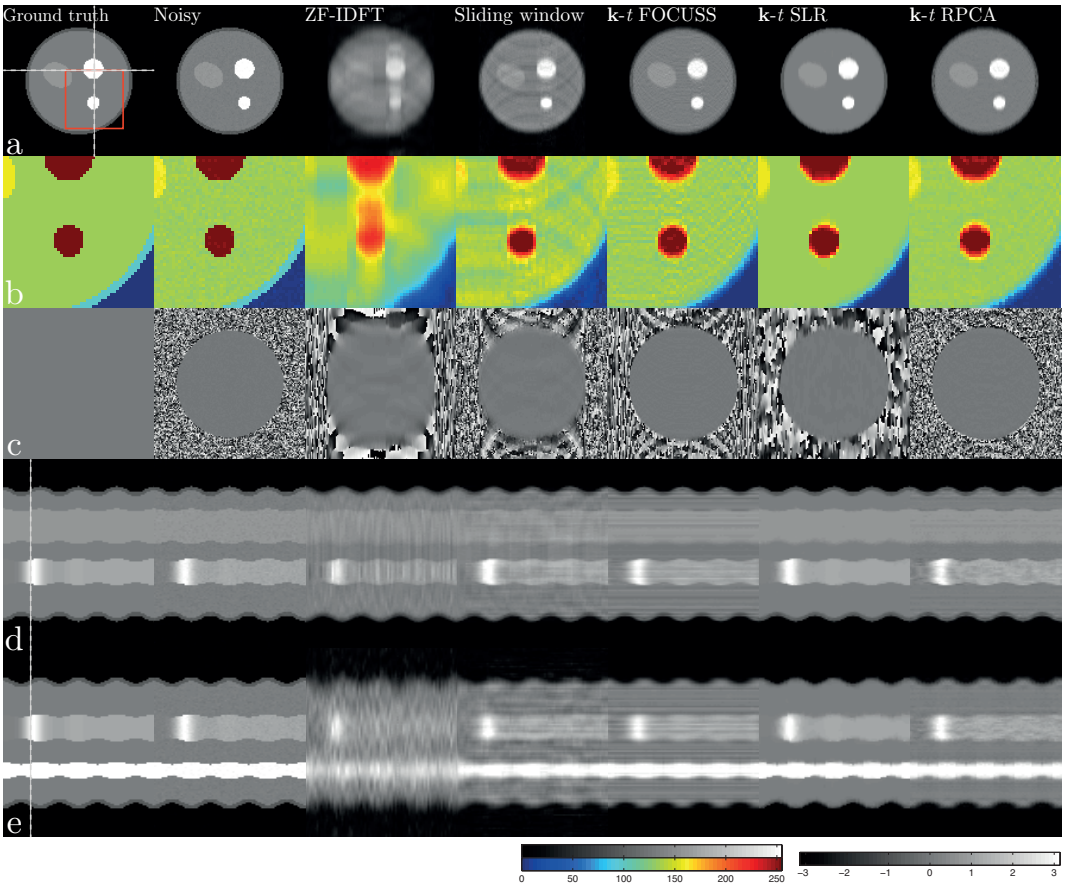

Figure 6.6: Qualitative results for phantom with a combination of intensity and motion (Cartesian sampling). (a) Magnitude images (b) Zoom-in magnitude images (corresponding to the red square on the ground truth image) (c) Phase images (d) x-t temporal profiles and (e) y-t temporal profiles (according to the dotted lines on the ground truth image). The time frames shown in the first three rows correspond to the frames selected on the dotted lines on the temporal profiles. Left color mappings refer to magnitude images, right color mapping refers to phase images.

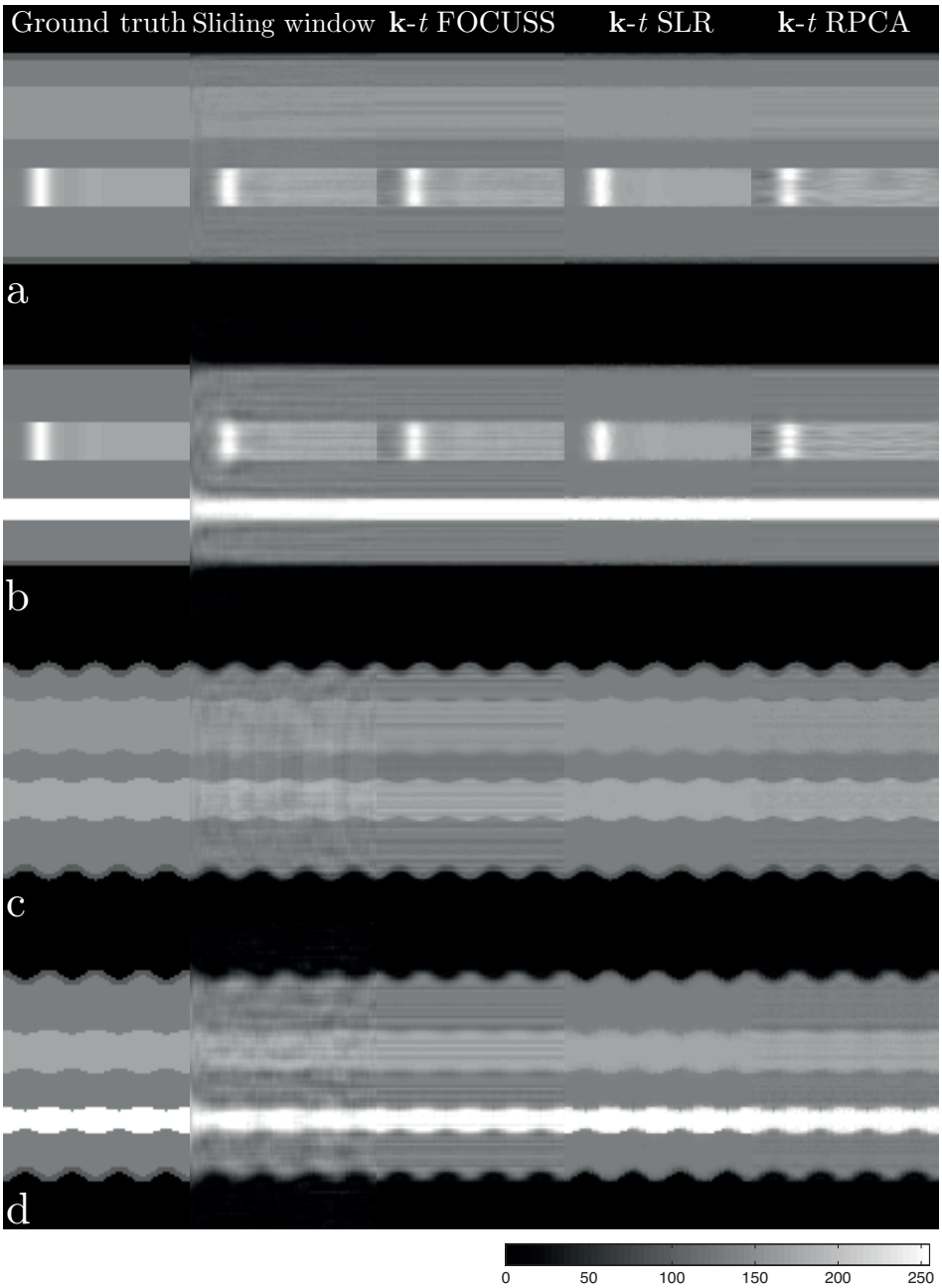

Figure 6.7: (a,c) x-t temporal profiles and (b,d) y-t temporal profiles of various reconstruction methods for (a,b) intensity only phantom and (c,d) motion only phantom (Cartesian sampling).

	Intensity only	Motion only	Combination
Rank	2	11	21
$\|\mathcal{F}_t(\mathbf{X})\|_1/\|\mathbf{X}\|_F$	79.7	96.5	98.9
$\mathrm{TV}(\mathbf{X})/\|\mathbf{X}\|_F$	47.4	60.1	60.7

Table 6.1: Characteristics of the different noiseless phantoms. The spatio-temporal TV operator is computed as defined in Eq. (3.62).

Phantom simulations

First experiments are conducted on a numerical phantom of dimensions $N_x = N_y = 128$, $N_t = 80$. This phantom is created to model typical dynamic MRI sequences with different types of time-varying components. Specifically, it can include periodic local and global motion, and localised changes of intensity. Local motion simulates moving organs (such as the beating heart) while global motion simulates respiratory-like movement imitating free-breathing imaging. Motion is modelled using trigonometric functions with varying frequencies and amplitudes. Local intensity changes mimic a contrast enhanced signal, i.e. the uptake and washout of a contrast agent using the modified Tofts model [147] as shown in figure 6.5. This is typical in dynamic contrast enhanced MRI studies. While simplistic, the major advantage of this phantom is the full control over motion and intensity parameters to make fine adjustments, and the availability of a reference noiseless signal (ground truth).

To evaluate the performance of the different reconstruction algorithms, the same phantom with different time-varying elements is used. Reconstruction methods are tested when the phantom has only intensity changes (no motion), periodic motion (no intensity changes) and with a combination of the two. Table 6.1 provides some characteristics of the three phantoms in the noiseless case, i.e. the rank, ℓ_1 norm of the Fourier transform along the time dimension and ℓ_1 norm of the gradient in x, y, t directions approximated by finite differences. Note that noiseless (ground truth) matrices are of dimensions $128^2 \times 80$ with various ranks depending on whether intensity/motion is present. However, noisy matrices are full-rank (80) because of the presence of noise, although they remain approximately low-rank with a number of significant singular values about equivalent to their noiseless counterpart.

In these experiments, the acceleration factor is approximately 10 (about 10% of acquired samples). Quantitative results are reported for Cartesian sampling and pseudo-radial sampling in tables 6.2 and 6.3. Reconstruction errors are shown in decibels and have been computed as in Eq. (B.2) with associated regularisation parameter(s) for the different methods in brackets. For **k**-t SLR, they refer to $[\alpha; \beta]$ and for **k**-t RPCA to $[\mu; \rho]$. Visual evaluations are provided in figures 6.6 and 6.7. Figure 6.6 shows magnitude images, phase images and temporal profiles for the phantom with a combination of intensity and motion. Figure 6.7 presents time

	Intensity only	Motion only	Combination
ZF-IDFT	14.5	14.6	14.5
Sliding window	21.3	16.9	16.9
k-t FOCUSS	**25.9** [0.1]	20.7 [0.1]	20.6 [0.1]
k-t SLR	25.7 [0;1]	**24.3** [1;1]	**24.2** [1;1]
k-t RPCA	24.5 [200;2]	21.7 [500;0.25]	21.5 [500;0.25]

Table 6.2: Reconstruction results (in dB) using Eq. (B.2) for numerical phantoms with Cartesian sampling. Numbers in brackets refer to regularisation parameters.

	Intensity only	Motion only	Combination
ZF-IDFT	15.8	15.9	15.9
Sliding window	23.8	17.3	17.3
k-t FOCUSS	31 [0.1]	22.7 [0.01]	22.6 [0.01]
k-t SLR	**29.9** [10;1]	**28.2** [10;1]	**28.2** [10;1]
k-t RPCA	28.1 [500;2]	23.8 [500;0.5]	23.6 [500;0.5]

Table 6.3: Reconstruction results (in dB) using Eq. (B.2) for numerical phantoms with pseudo-radial sampling. Numbers in brackets refer to regularisation parameters.

profiles of reconstructions for the phantom with only intensity and with only motion.

From these results, there is no indication in which **k**-t RPCA might have a preference for a certain type of dynamic phantom. In fact, it can be observed that **k**-t RPCA has a similar behaviour as **k**-t SLR, although **k**-t SLR consistently provides better reconstructions than **k**-t RPCA itself. Both **k**-t SLR and **k**-t RPCA outperform **k**-t FOCUSS when motion is present, and when a combination of intensity and motion is present. When only intensity is present however, **k**-t FOCUSS seems to have a slight advantage over both **k**-t SLR and **k**-t RPCA.

The general good performance of **k**-t SLR over other methods can possibly be attributed to the fact that the phantom is a piecewise constant signal. A spatio-temporal total variation prior is particularly efficient for this type of signal, since TV penalises oscillatory information while allowing jumps. In other words, this means that solutions obtained by **k**-t SLR are rather due to the sparsity prior than the low-rank one. For example, it can be seen that the low-rank structure is not exploited at all in the reconstruction of the phantom with only intensity for Cartesian sampling since α was selected equal to zero (table 6.2, "Intensity only" column).

Tables 6.2 and 6.3 show that all reconstruction methods benefit from the pseudo-radial sampling, whether they are based only on sparsity or both low-rank and sparsity prior. This could be be expected because it can be seen that a simple zero-filled inverse Fourier transform already gives an improved reconstruction performance when a pseudo-radial sampling pattern is employed in contrast to a Cartesian sampling pattern.

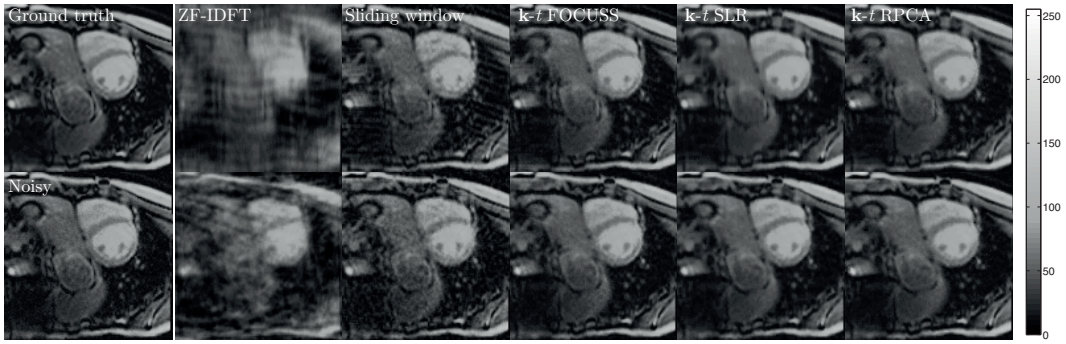

Figure 6.8: Visual comparison of reconstruction methods for cardiac MRI data showing one time frame magnitude image (frame number $n = 40$). First row corresponds to Cartesian sampling, second row to pseudo-radial sampling.

	Cartesian sampling	Pseudo-radial sampling
ZF-IDFT	9.5	11.0
Sliding window	13.8	14.9
k-t FOCUSS	17.2 [0.01]	18.5 [0.01]
k-t SLR	**17.8** [100;0.1]	19.2 [200;0.1]
k-t RPCA	17.5 [100;2]	**19.4** [100;2]

Table 6.4: Reconstruction results (in dB) using Eq. (B.2) for cardiac MRI data with Cartesian and pseudo-radial sampling.

Cardiac MRI

The second experiment is conducted on *in vivo* data with a free-breathing cardiovascular dataset of dimensions $N_x = N_y = 128$, $N_t = 90$ acquired on a 3T MRI scanner. Apart from motion such as heartbeats and large breathing movements, this dataset has complex anatomical features that makes it more challenging to reconstruct than the numerical phantom. An acceleration factor of approximately 8 is chosen, which corresponds to about 12.5% of acquired samples. Note that it is necessary to add noise to *in vivo* data because the original noise in the magnitude image becomes part of the apparent signal when retrospectively undersampled.

Time frames extracted from the different reconstruction methods are shown in

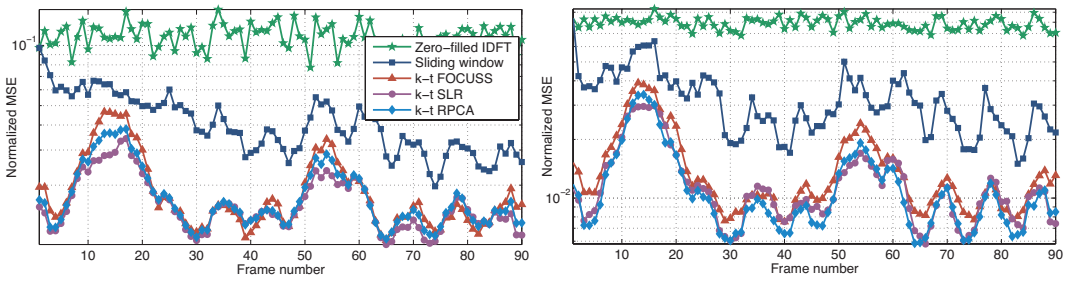

Figure 6.9: NMSE at each time frame for cardiac MRI data with Cartesian sampling (left) and pseudo-radial sampling (right).

figure 6.8, although it is visually difficult to claim objectively which method is the best. Quantitative results are given in table 6.4 for the different sampling strategies. Additionally, NMSE at each time frame are shown in figure 6.9 for both Cartesian and pseudo-radial sampling. Based on these results, all methods performed similarly using the polynomial variable density sampling, and a slight advantage can be seen for both **k**-t SLR and **k**-t RPCA when pseudo-radial sampling is used. In **k**-t RPCA, the selected reconstructions were chosen with $\rho = 2$ (table 6.4), which means that the selected reconstructions have favoured the low-rank part rather than the sparse part. **k**-t SLR did not successfully manage to obtain much better reconstruction results over other existing methods as it previously did on the first experiment. One of the possible reasons can be attributed to the fact that the cardiac MRI dataset did not have a particularly sparse gradient in space and time compared to the numerical phantoms.

6.3.3 Exploiting the separation

In this section, the utility of the intrinsic separation of the reconstructed data into low-rank and sparse components is demonstrated in the context of motion estimation in dynamic contrast enhanced (DCE) MRI.

In DCE MRI, acquisition of multiple MR images is taken continuously before, during, and after the administration of a contrast agent. The uptake and washout of the contrast agent concentration over time in the body corresponds to local changes of intensity in the MR images. Pharmacokinetic analysis can then be used to relate to tissue characteristics [147]. However, patient motion during acquisition (such as heartbeats, breathing or involuntary movements) produces inter-frame misalignment and complicates the estimation of the rate of the uptake by the tissue. Image registration can be used to solve this problem, but the presence of the contrast enhanced images interferes with the registration procedure because conventional algorithms can interpret local intensity changes as motion.

As it has been demonstrated for RPCA in figure 6.2, the proposed **k**-t RPCA approach is also expected to separate slow time-varying elements (background) from more abrupt changes due to the embedded separation. We show in figure 6.10 the different decompositions obtained with various ρ using the numerical phantom of the previous section that includes a combination of local intensity changes and slow varying motion. This figure demonstrates that it is possible to some degree to separate local changes of intensity in the sparse component from the background for an appropriate ρ parameter. To help in motion estimation, the idea is then to register low-rank images that will include most of the slow varying motion and less local intensity changes provoked by the contrast agent. The displacement field obtained from registering only the low-rank images (without interference from local intensity changes) is likely to be closer to the "ground truth displacement field",

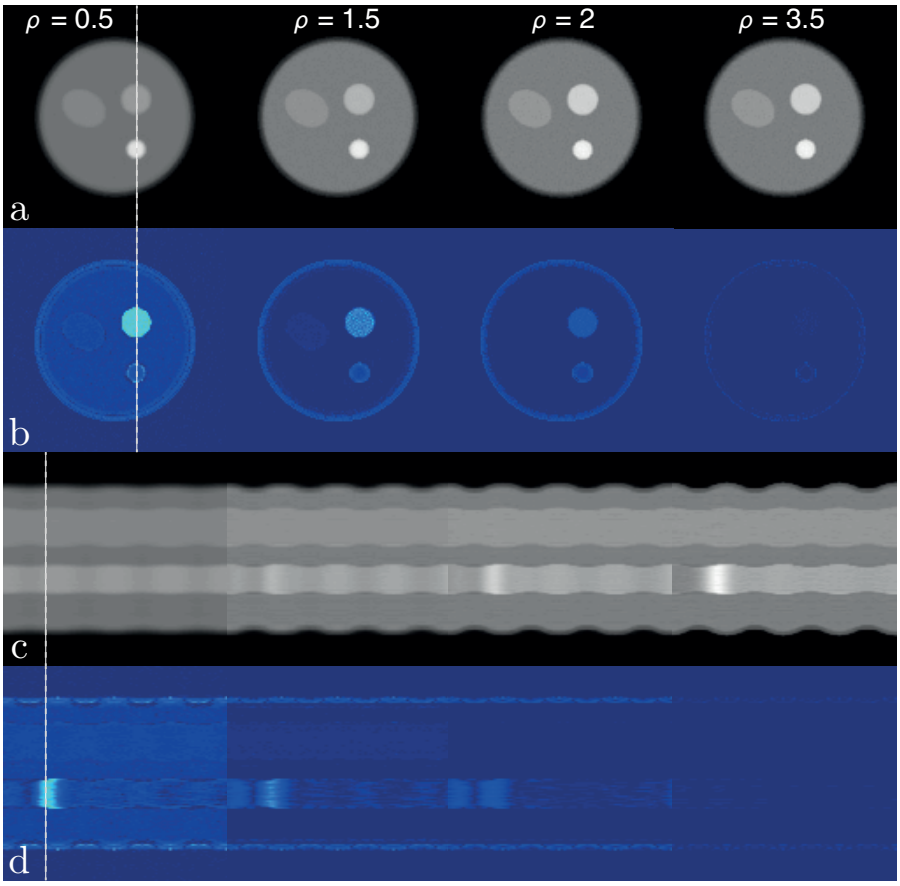

Figure 6.10: Different types of separation into low-rank and sparse components using **k**-t RPCA with different decomposition parameters ρ. It can be observed that this parameter acts as a trade-off between the two components. The undersampling rate is 0.25. (a) Low-rank time frames (b) Sparse time frames (c) x-t temporal profiles of low-rank component (d) x-t temporal profiles of sparse component.

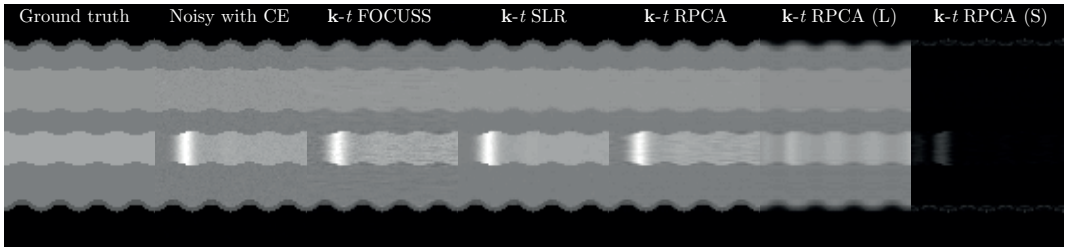

Figure 6.11: x-t temporal profiles used in the registration procedure. For the registration of **k**-t RPCA, only the low-rank part is used which mostly contains images without contrast enhancement (CE) thanks to the separation process.

that is the displacement field from the same MR signal without contrast enhanced images. This displacement field can then be employed for more accurate motion correction as it has been shown in Ref. [148].

As a proof of concept, the numerical phantom of the previous section (combination of local intensity changes and motion) is used for the purpose of demonstration with an acceleration factor of 4 using the pseudo-radial sampling. **k**-t FOCUSS and **k**-t SLR are reconstructed as previously, using the best regularisation parameters. However, **k**-t RPCA reconstruction is now explicitly selected with $\rho = 1.5$ to end up with mostly the local intensity changes in the sparse part, and motion in the low-rank part. Reconstruction errors for **k**-t FOCUSS, **k**-t SLR and **k**-t RPCA were respectively 25.3dB, 31.1dB and 26.3dB.

A sequential registration of each frame of **k**-t FOCUSS, **k**-t SLR and the low-rank part of **k**-t RPCA is performed with NiftyReg [149], an efficient C++ implementation of a parallel formulation of the free-form deformation (FFD) algorithm [150] based on cubic B-splines. Local normalised cross correlation is used as measure of similarity (standard deviation of the Gaussian kernel set to 5 pixels for all time points) and a control point spacing of 2 pixels in all directions. The time profiles of the different reconstruction methods along the ground truth are shown in figure 6.11. Note the ground truth was obtained with the noiseless phantom created without intensity changes but with the same motion. The reference images taken for registration were the last time frame images in the respective dynamic reconstructed sequences.

In the registration procedure, we also compute the displacement fields of the image sequence (which is called *optical flow* in computer vision) via NiftyReg. This results in displacement fields D_x and D_y, respectively along the x direction and y direction. Displacement vector fields from one time frame to the next time frame are shown (in blue) in figure 6.12 with source images used for registration in the background. The definition of the Jacobian $J(x, y)$ in 2D depends on the displacement fields D_x and D_y,

$$J(x,y) = \begin{bmatrix} \frac{\partial D_x}{\partial x} & \frac{\partial D_y}{\partial x} \\ \frac{\partial D_x}{\partial y} & \frac{\partial D_y}{\partial y}, \end{bmatrix} \tag{6.19}$$

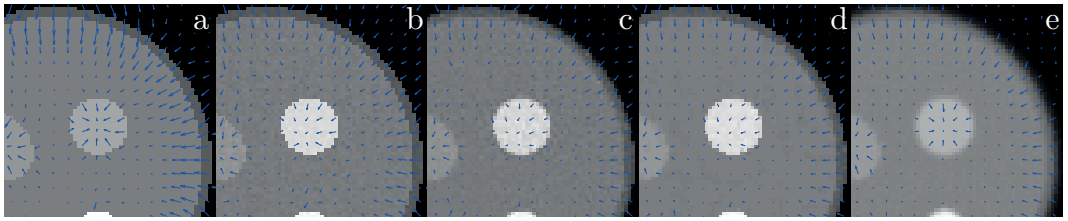

Figure 6.12: Displacement fields (zoom-in) over source images used for registration. Table 6.5 provides the associated quantitative results. (a) Ground truth noiseless phantom (b) Noisy phantom with local intensity changes (c) **k**-t FOCUSS (d) **k**-t SLR (e) **k**-t RPCA, low-rank part. It can be seen that the displacement field is better estimated in the region with local changes of intensity in **k**-t RPCA.

Noisy phantom with local intensity changes	11.0
k-t FOCUSS	10.4
k-t SLR	14.6
k-t RPCA – Low-rank component	**15.2**

Table 6.5: Displacement fields results in the region of interest with local intensity changes. Quantities are in dB and have been computed using the Jacobian and Eq. (B.2).

and in particular its determinant is given by

$$|J(x,y)| = \frac{\partial D_x}{\partial x}\frac{\partial D_y}{\partial y} - \frac{\partial D_x}{\partial y}\frac{\partial D_y}{\partial x}. \tag{6.20}$$

The metric defined in Eq. (B.2) can then be used to compute the displacement fields errors based on the Jacobian. Results are reported in table 6.5 and show a slight improvement for **k**-t RPCA over other methods.

6.4 Discussion

6.4.1 Prior assumptions and regularisation parameters

Experiments suggest that from a reconstruction point of view, the prior assumption made in **k**-t RPCA is strong enough to remain competitive with state of the art methods. The prior assumption in **k**-t FOCUSS is that the (x, f)-space is a sparse signal, which is appropriate with dynamic datasets that exhibit periodicity in time. **k**-t SLR performed well over other methods for phantom simulations, but was not so successful for the *in vivo* cardiac dataset.

Generally, regularisation parameters affect directly the reconstruction results, and selecting the right one is challenging in most inverse problems. In **k**-t FOCUSS and **k**-t SLR, the best reconstructions were selected by testing a range of different regularisation parameters. Two strategies for the choice of ρ in **k**-t RPCA have been adopted. If interested in obtaining the best reconstruction, the choice of ρ

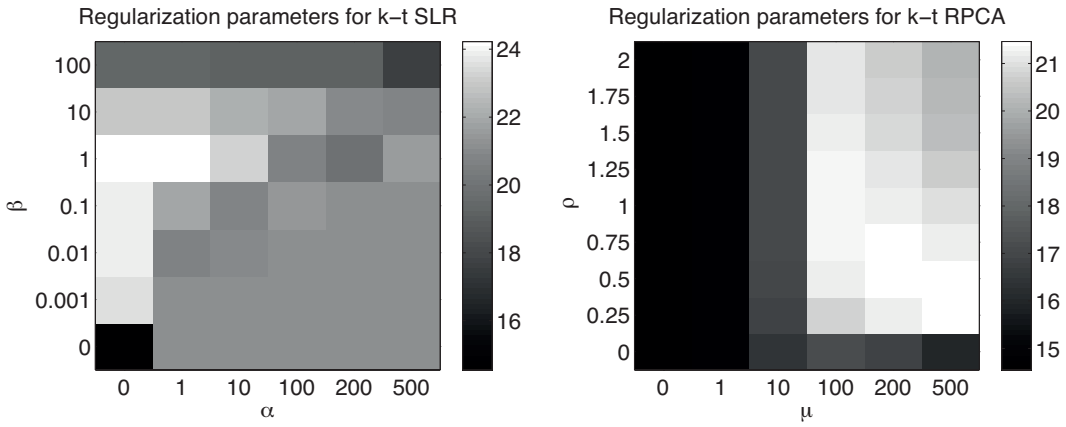

Figure 6.13: Influence of regularisation parameters on the reconstruction error (in dB) for **k**-t SLR and **k**-t RPCA. Numerical phantom simulations with 10-fold acceleration and Cartesian sampling. For **k**-t SLR, α and β refers respectively to the nonconvex Schatten norm and spatio-temporal gradient. For **k**-t RPCA, μ and ρ refers to (6.12).

should be selected such that the best reconstruction is returned. However, ρ can also be forced to a specific value to obtain a desirable decomposition. Figure 6.13 presents reconstruction errors (in dB) using different regularisation parameters for **k**-t SLR and **k**-t RPCA. This figure and previous experiments using the phantom suggest that the low-rank a priori information in **k**-t SLR is not playing an important role in this case. When $\alpha = 0$, it is observed that a good reconstruction can be obtained using only the sparsity a priori information (spatio-temporal TV). This can be attributed to the fact that piecewise constant signal such as the phantom is very sparse when the ℓ_1 norm of the gradient is computed. For **k**-t RPCA, the regularisation parameter μ is a trade-off between data consistency and the low-rank plus sparse decomposition. As shown in figure 6.13, the solution is not regularised when $\mu = 0$ and an appropriate value must be selected to obtain a good regularised solution. The decomposition parameter ρ can be varied, although it also affects the reconstruction results.

Note that the selection of regularisation parameters was optimised but this is an unrealistic strategy since the ground truth is not available in a practical scenario. However, methods can be adapted to find ideal regularisation parameters such as the discrepancy principle if noise properties are known, and for example L-curve or generalised cross-validation methods if not (see Ref. [151]).

6.4.2 Decomposition

While a separation into low-rank and sparse components is easy to see mathematically, it may be difficult to interpret physiologically what it represents in a dynamic MRI context. One of the reasons is that the decomposition depends on the type of dynamic data. Further work in this direction is needed to understand more deeply

how it can be interpreted physiologically.

Generally, the decomposition provides a separation into two components that have different characteristics. The low-rank component will tend to have slow time-varying elements while the sparse component will capture more abrupt changes, but ρ can be modified to balance between the two parts.

In the simulations, a specific example has been shown where partial isolation of local changes of intensity in the sparse component from the general motion leads to a better estimation of the displacement field. Departing from this example, it is likely that this combined reconstruction–separation approach offers further applications that could be investigated. For example, motion-related applications where sparse and localised motion elements interfere with the general background signal, or artefacts removal where the outlier component would cause undesired alterations of data. In the latter case, an artefact correction algorithm for RF spike noise has been proposed based on RPCA as a post-processing technique [152].

6.4.3 Noise

This study has included complex-valued noise to simulate more realistic experiments. However, this study has not evaluated the reconstruction (and separation) performance of the proposed approach as a function of added noise. Generally in **k**-t RPCA, increased noise is inclined to interfere with the sparse component. Both **k**-t FOCUSS and **k**-t SLR may be relatively more robust to noise regarding reconstructed data, since in **k**-t FOCUSS the noise in (x, f)-space will generally not be represented by highly sparse coefficients, and the spatio-temporal total variation norm will tend to smooth a noisy solution in **k**-t SLR.

6.4.4 Acquisition and sampling

Two strategies to undersample the $(\mathbf{k}, t)$-space, a Cartesian and pseudo-radial sampling patterns, have been used. These were respectively based on polynomial variable density and random rotations across each acquisition frame to produce incoherent sampling artefacts. From the experiment section, in particular tables 6.2, 6.3 and 6.4, it is suggested that the pseudo-radial undersampling strategy provides better reconstruction results for all methods either based on only sparsity as **k**-t FOCUSS, or based on low-rank and sparse prior information (**k**-t SLR, **k**-t RPCA). However, these sampling strategies were based on retrospectively undersampling the $(\mathbf{k}, t)$-space and they would ideally need to be validated using prospectively undersampled data from an MRI scanner.

6.4.5 Computational times

The purpose of this paper was not to focus on computational aspects of the different reconstruction methods. The different algorithms used were implemented in Matlab but not optimised. However, it should be noted that during our simulations, **k**-t FOCUSS reconstructions could be obtained in less than a minute, whereas **k**-t SLR and **k**-t RPCA could require several minutes of computation in contrast ($\simeq 10$ minutes).

6.5 Related works

The proposed method was inspired by using initially the L+S model for dynamic MRI as a post-processing technique [153]. Gao et al. have applied a similar approach for dynamic computed tomography [154], cardiac cine MRI [133, 155, 156] and diffusion MRI [157] where low-rank plus sparsity is used as a prior, but they do not emphasise or show the potential role of the separation.

In contrast, recent work by Otazo et al. [89, 158] have also highlighted the role of the low-rank plus sparse separation. In fact, they have published almost simultaneously the same model in Ref. [89]. Some key differences are as follows: (i) they combine the technique with parallel dynamic MRI, (ii) their algorithm is based on proximal gradient instead of ADMM in our case, (iii) they provide more clinical examples with retrospective and true undersampling whereas we have focused only on retrospective undersampling, (iv) the separation is shown to be useful in the case of time-resolved angiography (background suppression) and DCE MRI (separation of contrast-enhanced information from non-enhanced background); in our case we have shown a potential application for motion estimation of DCE MRI sequences, (v) they provide comparisons with standard CS reconstruction and simultaneous low-rank and sparsity only by modifying the manner in which the constraints are enforced in their optimisation algorithm, whereas we have provided comparisons with two state of the art techniques using the original algorithms of **k**-t FOCUSS and **k**-t SLR.

6.6 Conclusion

In this chapter, we have presented a method termed **k**-t RPCA that jointly reconstructs and separates dynamic MR data from partial measurements by employing a low-rank plus sparse regularisation prior. While providing a competitive reconstruction method for accelerated dynamic MRI as the comparison with state of the art methods have shown, this technique also provides a separation into two components having different characteristics that can have potential when tailored to the right

application. Here, the decomposition was used to partially separate the contrast enhanced region in a simulated DCE sequence and help in motion estimation.

Interestingly, the proposed reconstruction–separation approach suggests a method that is not only about finding the closest representation of the true object, but is also a step towards methods that would directly infer relevant characteristics from limited measurements.

Chapter 7

Low-rank based recovery for dynamic parallel imaging

Contents

7.1 Introduction

In this chapter, we investigate the combination of dynamic parallel MR imaging with methods based on low-rank structure.

Parallel MR imaging (PMRI) uses multiple coils that receive data simultaneously (i.e. in parallel) to reconstruct a signal from undersampled **k**-space data by using information about coil sensitivities. PMRI, also referred to as *multichannel* or *multicoil* MRI, is the most used technique in clinical practice to obtain accelerated scan times. PMRI has been combined with complementary acceleration approaches such as compressed sensing because both techniques reduce sampling based on different types of information, i.e. sparseness for CS and coil sensitivities for PMRI.

The contribution of this work is to further investigate low-rank based recovery methods in the context of dynamic parallel MR imaging. We mainly focus on low-rank promotion due to the novelty of this prior in PMRI and because of the overall good reconstruction performances that were obtained using low-rank regularisation alone in chapter 5. We develop and characterise low-rank based methods for PMRI such as low-rank regularised SENSE, coil-by-coil low-rank reconstruction, and the low-rank tensor approach as recently suggested by Trzasko and Manduca [159] to exploit the low-rank property in multiple dimensions of the data. Formulated convex optimisation problems are solved with proximal gradient algorithms. Content of this chapter has not been published.

This chapter is organised as follows. We recall some fundamental principles of PMRI and describe the proposed low-rank based methods for dynamic PMRI in section 7.2. Numerical simulations are presented in section 7.3. The discussion, related works and conclusion are presented in sections 7.4, 7.5 and 7.6.

7.2 Method

7.2.1 Dynamic parallel MRI

The use of multiple detectors in NMR (originally called phased array imaging) can be traced back to the late 1980s [160] and early 1990s [161] but it was only in the late 1990s that it was first used successfully to accelerate MRI.

The idea behind PMRI is to reconstruct a signal from undersampled **k**-space data by using multiple coils that receive data in parallel. Each coil provides full FOV images but with artefacts because of the sub-Nyquist sampling. However, the MR image can be reconstructed from multichannel **k**-space data sampled at the sub-Nyquist rate due to the availability of multiple independent receiver coils with distinct sensitivities. In theory, the acceleration factor can be up to the number of coils, but in practice it is limited by noise and imperfect coil geometry.

The aim of PMRI is to reconstruct images by exploiting coil sensitivity profiles and sub-Nyquist **k**-space samples. Since this step can be performed either directly in **k**-space or in the image domain, PMRI methods have been traditionally categorised into Fourier domain (**k**-space) and image domain approaches depending on the data representation used to perform the reconstruction: SMASH [162] and GRAPPA [163] are Fourier domain based, whereas SENSE [164] is image domain based. The various PMRI methods also differ in the manner in which they use sensitivity information: SMASH and SENSE assume that the sensitivity profiles from each coil are known while GRAPPA does not. In practice, sensitivity calibration is achieved by a pre-scan but it may be difficult to obtain sensitivity profiles with high accuracy due to motion. When techniques do not need to know them, they are referred to as *autocalibrating*

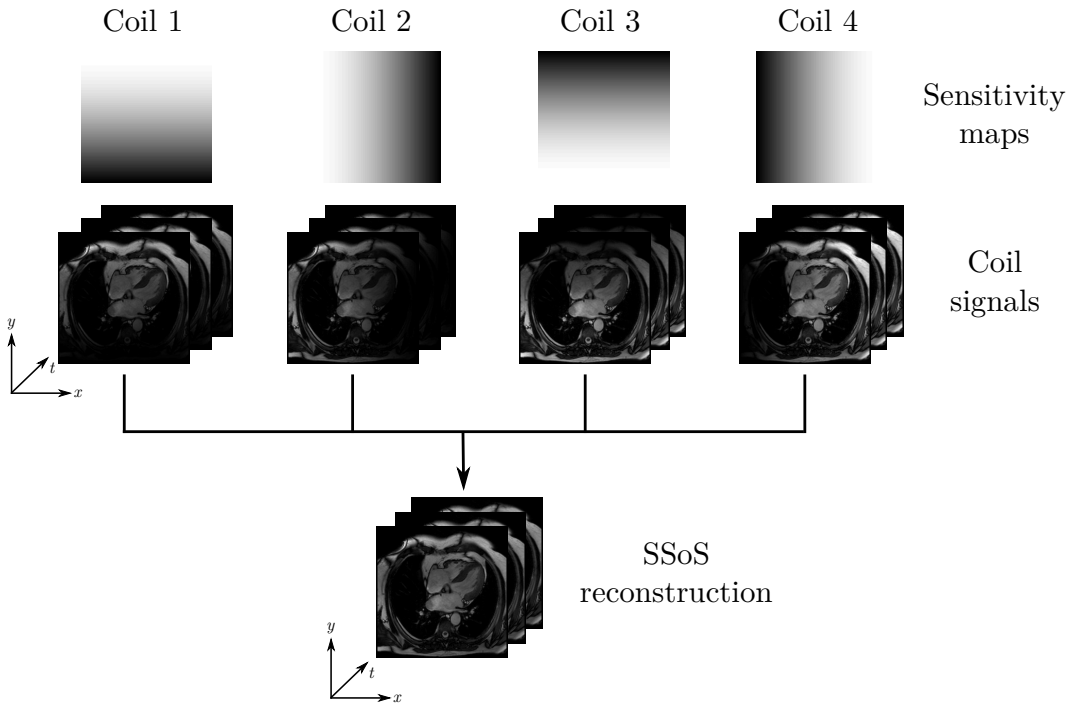

Figure 7.1: The concept of dynamic parallel MRI illustrated with 4 coils ($N_c = 4$) in the image domain (magnitude). For illustration purpose, sampling is at the Nyquist rate here but in practice it would be at sub-Nyquist rate and images would show artefacts. Each coil signal exhibit high correlation in time. SSoS stands for square root of sum of squares.

or *calibrationless* methods. In clinical practice, SENSE and GRAPPA are the two most popular PMRI methods.

The parallel imaging technique for dynamic MRI is illustrated with 4 coils in figure 7.1. In practice, it is possible to have access to coil arrays with much more than 4 channels (such as 64). Standard methods to reconstruct data from multiple coil signals can be done in various ways, such as simple complex sum, sum of squares, sensitivity-weighted sum, or square root of sum of squares (SSoS). More formally, the multicoil spatio-temporal imaging equation can be expressed as

$$S_j(\mathbf{k}, t) = \int_{\mathbb{R}^D} I(\mathbf{r}, t)\Gamma_j(\mathbf{r}, t)e^{-i2\pi(\mathbf{r}\cdot\mathbf{k})}d\mathbf{r} + N_j(\mathbf{k}, t), \tag{7.1}$$

where S_j refers to the $(\mathbf{k}, t)$-space signal for coil j with sensitivity profile Γ_j, I refers to the spatio-temporal image function, N_j is the spatio-temporal coil-dependent Gaussian noise function and $\mathbf{r}$, $\mathbf{k}$, t represent respectively the coordinates of spatial positions, the coordinates of spatial frequencies, and the time variable.

7.2.2 Sensitivity encoding reconstructions

The equivalent finite-dimensional model of Eq. (7.1) can be written

$$\mathbf{y} = \mathcal{E}_\Gamma(\mathbf{x}) + \mathbf{n}. \tag{7.2}$$

In this model, $\mathcal{E}_\Gamma : \mathbb{C}^{N_x N_y N_t} \to \mathbb{C}^M$ represents the parallel MRI encoding operator modelling sub-Nyquist sampling (not necessary random here), Fourier transform and coil sensitivities Γ_j, $\mathbf{y} \in \mathbb{C}^M$ is the stacked $(\mathbf{k}, t)$-space measurements vector for each coil and $\mathbf{n} \in \mathbb{C}^M$ is the stacked normally distributed noise vector for each coil. Eq. (7.2) represents the standard sensitivity encoding (SENSE) MRI model [164] within a dynamic imaging context.

k-t SENSE

The reconstruction of dynamic PMRI data using coil sensitivities information can be done by solving

$$\min_{\mathbf{x} \in \mathbb{C}^{N_x N_y N_t}} \frac{1}{2} \|\mathcal{E}_\Gamma(\mathbf{x}) - \mathbf{y}\|_2^2, \tag{7.3}$$

which is equivalent in the literature to **k**-t SENSE [71]. Although problem (7.3) has a closed form solution, in practice it is generally computed through iterative methods and not explicitly.

Regularised k-t SENSE

A notable work combining dynamic PMRI and compressed sensing (sparsity) is **k**-t SPARSE-SENSE proposed by Otazo et al. [165]. This method considers the **k**-t SENSE problem (7.3) with an additional sparsity prior. The optimisation problem considered is

$$\min_{\mathbf{x} \in \mathbb{C}^{N_x N_y N_t}} \frac{1}{2} \|\mathcal{E}_\Gamma(\mathbf{x}) - \mathbf{y}\|_2^2 + \alpha \|\mathcal{F}_t(\mathbf{x})\|_1, \tag{7.4}$$

where in this case $\mathcal{E}_\Gamma$ models *random* undersampling and $\mathcal{F}_t$ is the temporal Fourier transform. The combination of PMRI and CS has been shown to improve reconstruction results compared to using these techniques alone.

In this work, one of the contributions is to also consider **k**-t low-rank SENSE, i.e. the minimisation of the low-rank regularised **k**-t SENSE to exploit both low-rank a priori and coil sensitivities,

$$\min_{\mathbf{X} \in \mathbb{C}^{N_x N_y \times N_t}} \frac{1}{2} \|\mathcal{E}_\Gamma(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha \|\mathbf{X}\|_*. \tag{7.5}$$

There is no closed form solution for both problems (7.4) and (7.5), but minimisations of these functionals are possible under the proximal splitting framework and in particular using the fast proximal gradient algorithms.

7.2.3 Calibrationless approaches

We now describe other methods to reconstruct dynamic parallel MRI data without the need of incorporating coil sensitivity maps into the model. These methods can be considered calibrationless because they do not require coil sensitivity calibration and they only exploit coil sensitivities implicitly.

Coil-by-coil reconstructions

A basic method to reconstruct data without modelling coil sensitivities in the model would be to promote sparsity structure for each coil. This can be done by considering these N_c minimisation problems,

$$\min_{\mathbf{x}_j \in \mathbb{C}^{N_x N_y N_t}} \Big\{ F(\mathbf{x}_j) \equiv \frac{1}{2} \|\mathcal{E}(\mathbf{x}_j) - \mathbf{y}_j\|_2^2 + \alpha \|\Psi(\mathbf{x}_j)\|_1 \Big\}, \quad \forall j = 1, \dots, N_c. \tag{7.6}$$

Once a solution vector $\widehat{\mathbf{x}}_j$ has been found for each coil j, the single dynamic sequence can be obtained using a standard method to reconstruct data from multiple channels such as SSoS.

Similarly, each coil signal can be seen as a $N_x N_y \times N_t$ Casorati matrix which is very likely, as in single coil dynamic MRI, to have only few significant singular values due to high correlation in time (see figure 7.1). Thus, low-rank structure can also be promoted for each coil by considering the N_c minimisation problems

$$\min_{\mathbf{X}_j \in \mathbb{C}^{N_x N_y \times N_t}} \Big\{ F(\mathbf{X}_j) \equiv \frac{1}{2} \|\mathcal{E}(\mathbf{X}_j) - \mathbf{y}_j\|_2^2 + \alpha \|\mathbf{X}_j\|_* \Big\}, \quad \forall j = 1, \dots, N_c, \tag{7.7}$$

where $\mathbf{X}_j \in \mathbb{C}^{N_x N_y \times N_t}$ represents the Casorati matrix for coil j to recover.

Note that in both problems (7.6) and (7.7), the MRI encoding operator $\mathcal{E}$ does not model coil sensitivity maps. These methods are respectively referred to as "coil-by-coil sparse" and "coil-by-coil low-rank" hereinafter.

Low-rank tensor reconstruction

Another calibrationless method is to consider dynamic parallel MRI within a low-rank tensor framework which was suggested in a short abstract by Trzasko and Manduca [159]. Low-rank tensor recovery can be seen as an extension of low-rank matrix recovery as it has been explained in section 3.3.5. The low-rank tensor framework enables the exploitation of low-rank structure in multiple dimensions of the data, which is generally desirable from an inverse problem point of view since additional a priori information is used.

Trzasko and Manduca [159] have suggested that dynamic PMRI data could be viewed as a tensor $\boldsymbol{\mathcal{X}}$ of order $Q = 3$ of the form "Space $\times$ Time $\times$ Coils", or more formally as $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{N_{xy} \times N_t \times N_c}$ where we adopt the notation $N_{xy} = N_x N_y$. The

Notation	Dimensions	Meaning
$\boldsymbol{\mathcal{X}}$	$N_{xy} \times N_t \times N_c$	Space × Time × Coils
$\mathbf{X}_{(1)}$	$N_{xy} \times N_t N_c$	Space × Time-Coils
$\mathbf{X}_{(2)}$	$N_t \times N_{xy} N_c$	Time × Space-Coils
$\mathbf{X}_{(3)}$	$N_c \times N_{xy} N_t$	Coils × Space-Time

Table 7.1: The third-order tensor $\boldsymbol{\mathcal{X}}$ and its mode-q unfolding $\mathbf{X}_{(q)}$.

tensor-based forward MRI model can be written

$$\mathbf{y} = \mathcal{E}(\boldsymbol{\mathcal{X}}) + \mathbf{n}, \tag{7.8}$$

where $\mathbf{y} \in \mathbb{C}^M$ is the stacked $(\mathbf{k}, t)$-space measurements vector for each coil, $\mathcal{E} : \mathbb{C}^{N_{xy} \times N_t \times N_c} \to \mathbb{C}^M$ represents the MRI encoding operator modelling the random sub-Nyquist sampling ($M \ll N_{xy} N_c N_t$) and Fourier transform for each coil in the spatial tensor dimension only, and $\mathbf{n} \in \mathbb{C}^M$ is the normally distributed noise vector.

The tensor $\boldsymbol{\mathcal{X}}$ can be unfolded in the different modes as shown in table 7.1. Unfolding a tensor of order Q is also called matricisation because it results in Q matrices, as it was explained in more details in section 3.3.5. Figure 7.2 shows the singular values of the different unfolding modes for a typical dynamic parallel MRI dataset of dimensions $N_x N_y = 128^2, N_t = 40, N_c = 16$. Table 7.1 and figure 7.2 highlight the fact that the tensor framework for dynamic PMRI can be seen as a generalisation of the Casorati matrix that has been considered so far. When only one coil is considered ($N_c = 1$), the mode-1 unfolding matrix $\mathbf{X}_{(1)}$ is equivalent to the Casorati matrix defined as in Eq. (3.59) for single coil time-series images.

To exploit the rank deficiency in the multiple unfolding modes, one can naturally consider the low-rank tensor recovery problem that uses the convex q-rank as in Eq. (3.47),

$$\boldsymbol{\mathcal{X}}^\star = \arg\min_{\boldsymbol{\mathcal{X}}} \Big\{ F(\boldsymbol{\mathcal{X}}) \equiv \frac{1}{2} \|\mathcal{E}(\boldsymbol{\mathcal{X}}) - \mathbf{y}\|_2^2 + \sum_{q=1}^{Q} \alpha_q \|\mathbf{X}_{(q)}\|_* \Big\}. \tag{7.9}$$

In this minimisation problem, the low-rank property is assumed jointly in all unfolding modes of the tensor, but the parameters α_q's can be individually tuned in order to provide regularisation weights to the nuclear norms. This method is referred to as "low-rank tensor" hereinafter.

We introduce the fast proximal gradient algorithm 7.1 to solve the unconstrained problem (7.9), inspired by algorithms proposed in Refs. [66, 114]. The principal step in this algorithm is the singular value soft thresholding of the mode-q unfolding of the tensor. The unfolding operation UnFold(.) consists in obtaining the different matrices from the tensor, while the folding operation Fold(.) consists in reconstructing the original tensor from the different matrices. To lighten the notation, we use

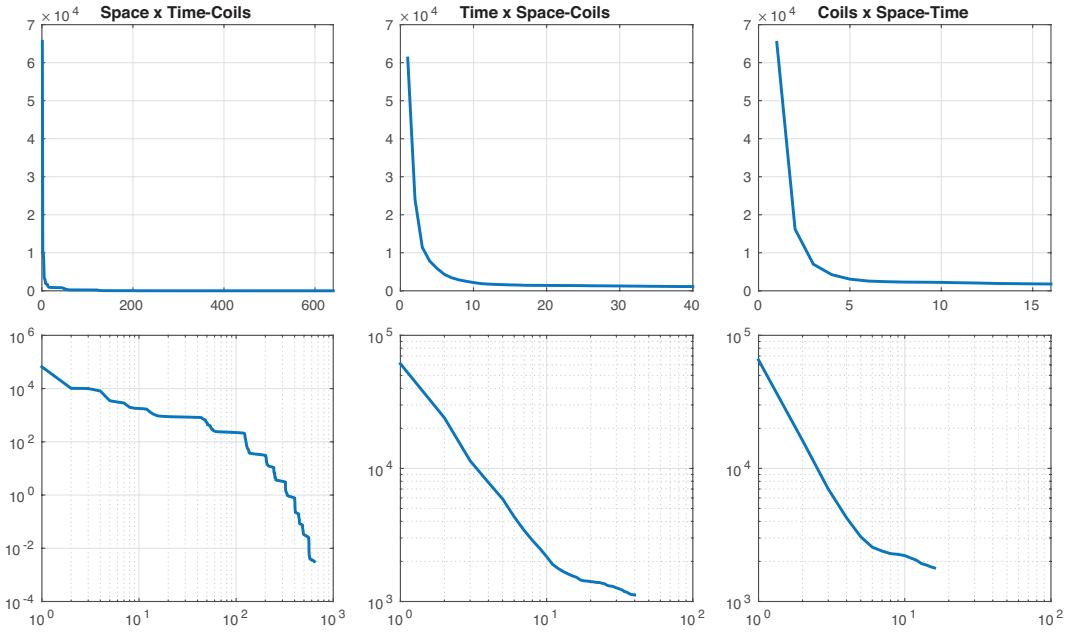

Figure 7.2: Singular values of the unfolding modes $\mathbf{X}_{(q)}$. These figures were generated with a signal of dimensions $N_x N_y = 128^2, N_t = 40, N_c = 16$. All unfolding modes have only a few significant singular values, although to different levels. Second row shows graphs in logarithmic scale.

Algorithm 7.1 Low-rank tensor recovery for dynamic parallel MR imaging via fast proximal gradient

Input: $\mathbf{y}$, $\alpha_q \geq 0$, $\rho > 0$
Initialise: $k = 0$, $\boldsymbol{\mathcal{X}}^0 = \boldsymbol{\mathcal{W}}^0 = \mathbf{0}$
while stopping criterion is not met **do**
  $\mathbf{W}_{(q)}^k \leftarrow \text{UnFold}(\boldsymbol{\mathcal{W}}^k)$
  for $q = 1, \ldots, Q$ **do**
    $\mathbf{X}_{(q)}^{k+1} \leftarrow \text{SVT}^{\mathcal{S}_{\rho\alpha_q}}(\mathbf{W}_{(q)}^k - \rho\mathcal{E}^*(\mathcal{E}(\mathbf{W}_{(q)}^k) - \mathbf{y}))$
  end for
  $\boldsymbol{\mathcal{X}}^{k+1} \leftarrow \frac{1}{Q}\sum_q^Q \text{Fold}(\mathbf{X}_{(q)}^{k+1})$
  $\boldsymbol{\mathcal{W}}^{k+1} \leftarrow \boldsymbol{\mathcal{X}}^{k+1} + \frac{k-2}{k+1}(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k)$
end while
Output: $\widehat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{X}}^k$

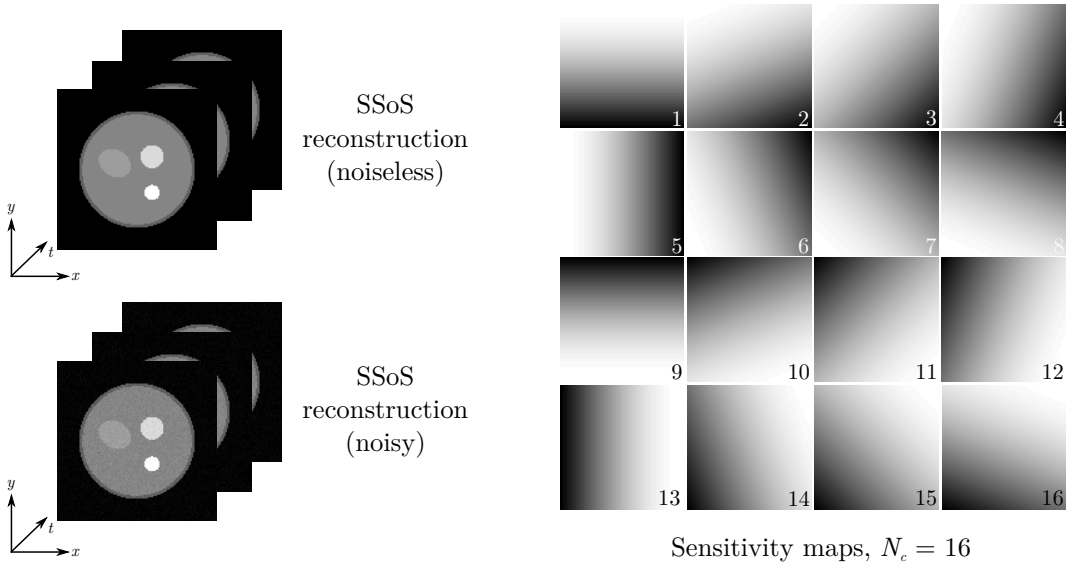

Figure 7.3: The dynamic PMRI phantom used for numerical simulations with artificially simulated sensitivity maps. Dimensions are $N_x = N_y = 128, N_t = 40, N_c = 16$.

the same MRI encoding operators $\mathcal{E}, \mathcal{E}^*$ for the different unfolding modes.

7.3 Numerical simulations

7.3.1 Framework

A synthetic, complex-valued, numerical phantom is used with simulated coil sensitivity maps. We use a numerical phantom based on the Shepp-Logan phantom with a combination of local and global motion (to simulate free-breathing data) and with local intensity changes. The dimensions of the phantom are $N_x = N_y = 128, N_t = 40, N_c = 16$. Spatially smooth coil sensitivity profiles are generated artificially. This enables us to have access to a noiseless reference signal as well as coil sensitivities profiles without doing a pre-scan for sensitivity calibration. The dataset along with coil sensitivity profiles are shown in figure 7.3.

We use retrospectively radial-based (equispaced angle) undersampling with an acceleration factor of 8 due to the superior performance of this sampling strategy for both sparse signal and low-rank matrix recovery as shown by results reported in section 5.3.2. Intensity of the phantom is normalised between values 0 and 255 in magnitude prior to any processing. Gaussian noise is added explicitly on each real and imaginary channel with a standard deviation $\sigma = 5$. The set of regularisation parameters tested is $\{0, 10^k : k \in \mathbb{Z}; -4 \leq k \leq 3\}$. For all algorithms, the step size is chosen as $\rho = 1$ and they are stopped if a maximum number of 100 iterations is reached, or if $F(\mathbf{x}^{k+1}) - F(\mathbf{x}^k)/F(\mathbf{x}^k)$ is less or equal to 10^{-5}.

Method	Reconstruction performance	Regularisation parameter(s)
ZF-IDFT	17.0	-
k-t SENSE [71]	20.8	-
k-t SPARSE-SENSE [165]	22.0	$\alpha = 0.01$
k-t low-rank SENSE	21.8	$\alpha = 10$
Coil-by-coil sparse	22.0	$\alpha = 0.01$
Coil-by-coil low-rank	20.6	$\alpha = 1$
Low-rank tensor	18.7	$\alpha_{1,2,3} = 10$
Low-rank tensor	20.4	$\alpha_1 = 10, \alpha_{2,3} = 0$
Low-rank tensor	16.7	$\alpha_2 = 10, \alpha_{1,3} = 0$
Low-rank tensor	16.7	$\alpha_3 = 10, \alpha_{1,2} = 0$

Table 7.2: Reconstruction results (in dB) using Eq. (B.2). Note that **k**-t SENSE, **k**-t SPARSE-SENSE and **k**-t low-rank SENSE require an estimation of the coil sensitivity profiles, in contrast to coil-by-coil sparse, coil-by-coil low-rank and low-rank tensor reconstructions that make use of implicit coil sensitivities. The notation $\alpha_{1,2,3} = 10$ means that α_1, α_2 and α_3 are all set to 10.

7.3.2 Reconstruction results

The following reconstruction methods are evaluated:

- **k**-t SENSE [Eq. (7.3)],
- **k**-t SPARSE-SENSE [Eq. (7.4)],
- **k**-t low-rank SENSE [Eq. (7.5)],
- coil-by-coil sparse [Eq. (7.6)],
- coil-by-coil low-rank [Eq. (7.7)],
- low-rank tensor [Eq. (7.9)].

The **k**-t SENSE reconstruction is computed with either **k**-t SPARSE-SENSE or **k**-t low-rank SENSE by setting regularisation parameter α to zero. Reconstruction results (in dB) are reported in table 7.2 with the regularisation parameters that provided the best reconstructions.

First, these results confirm that sparsity regularised SENSE (**k**-t SPARSE-SENSE) and low-rank regularised SENSE (**k**-t low-rank SENS) reconstructions perform better than SENSE alone. This is because PMRI and low-dimensional signal recovery methods are complementary acceleration approaches to each other. A slight advantage can be observed for the sparsity-based regularisation.

The good performance of coil-by-coil reconstructions (and in particular coil-by-coil sparse) may be explained by the nature of the dataset and coil sensitivity maps used in these simulations. Indeed, the numerical phantom possesses both global

and local periodic motions, which is particularly adequate for the temporal Fourier transform. Additionally, the coil sensitivity maps that were generated artificially can be qualified as simplistic as shown in figure 7.3.

Finally, observe that we have computed different low-rank tensor reconstructions by varying the regularisation parameters α_q in Eq. (7.9) in order to understand which unfolding has more importance. As it is suggested by the results, not all unfolding modes are actually helpful, since surprisingly the best reconstruction is obtained when both α_2 and α_3 are set to zero. While figure 7.2 shows that all unfolding modes have only a few significant singular values, not all unfolding matrices $\mathbf{X}_{(q)}$ can be qualified as low-rank. The reason is that one dimension of the unfolding matrix is too small compared to the other dimension. For instance, if we consider the matrix $\mathbf{X}_{(3)} \in \mathbb{C}^{N_c \times N_{xy} N_t}$, we know that

$$\operatorname{rank}(\mathbf{X}_{(3)}) \leq \min(N_c, N_{xy} N_t). \tag{7.10}$$

In general, we expect to have $N_{xy} N_t > N_c$ in a typical dynamic PMRI scenario, so that $\operatorname{rank}(\mathbf{X}_{(3)}) \leq N_c$. The matrix $\mathbf{X}_{(3)}$ may have only a few significant singular values, but it is likely that it cannot be qualified as low-rank simply because

$$\operatorname{rank}(\mathbf{X}_{(3)}) \leq N_c \ll N_{xy} N_t. \tag{7.11}$$

This emphasises the fact that there is an important distinction to make for a matrix in having only a few significant singular values and being characterised as low-rank.

7.3.3 Local low-rank tensor reconstruction results

The limitation of the low-rank tensor approach can be overcome by promoting low-rank locally instead of globally in the image domain as suggested by Trzasko and Manduca [142, 159]. We have shown the clear advantage of this approach in terms of reconstruction results in section 5.3.6 for single coil undersampled dynamic MRI. To promote low-rank locally within the tensor framework, only a slight modification of the problem (7.9) is needed, which can be written

$$\boldsymbol{\mathcal{X}}^\star = \arg\min_{\boldsymbol{\mathcal{X}}} \Big\{ F(\boldsymbol{\mathcal{X}}) \equiv \frac{1}{2} \|\mathcal{E}(\boldsymbol{\mathcal{X}}) - \mathbf{y}\|_2^2 + \sum_{q=1}^{Q} \alpha_q \sum_{\omega \in \Omega} \|\mathcal{R}_\omega(\mathbf{X}_{(q)})\|_* \Big\}, \tag{7.12}$$

where $\mathcal{R}_\omega$ is an operator that extracts local blocks of the Casorati matrix of the mode-q unfolding $\mathbf{X}_{(q)}$ and Ω represents the set of all blocks. Algorithm 7.1 is modified to compute blockwise soft thresholding instead of global soft thresholding for the different unfolding modes.

Reconstruction results are summarised in table 7.3 where we have tested different sets of nonoverlapping blocks. Note the improvement of reconstruction results as

Ω	Reconstruction performance (dB)	Reconstruction time (min)
1	18.7	32.4
4	18.9	26.3
16	20.2	25.6
64	21.4	11.6

Table 7.3: Reconstruction results (in dB) using Eq. (B.2) for locally low-rank tensor recovery with $\alpha_1 = \alpha_2 = \alpha_3 = 10$ and different Ω.

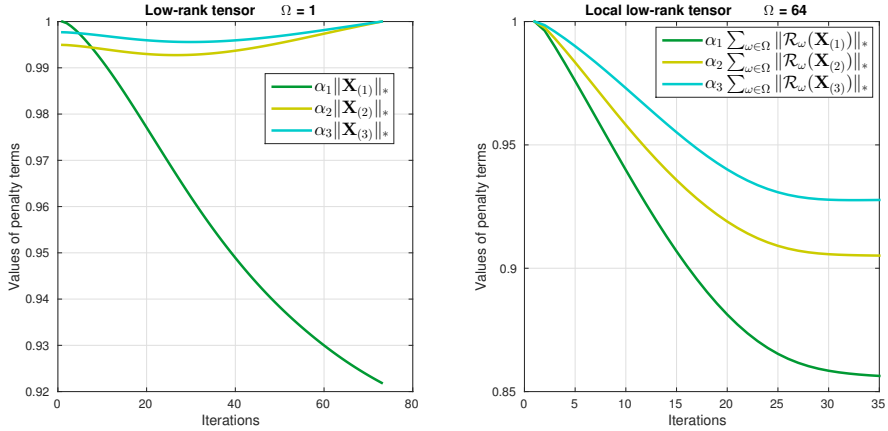

Figure 7.4: Values of penalty terms (normalised) in the objective function (7.9) (low-rank tensor, left figure) and in the objective function (7.12) (local low-rank tensor with $\Omega = 64$, right figure).

the set of blocks Ω is increased. Also, observe the decrease in reconstruction times as the number of blocks Ω increases, which is due to performing SVD on smaller matrices.

In figure 7.4, we illustrate the behaviours of the penalty terms for low-rank tensor ($\Omega = 1$) and local low-rank tensor with a number of blocks $\Omega = 64$. More specifically, the left figure shows the value per iterations of the penalty terms for low-rank tensor, i.e. $\alpha_1\|\mathbf{X}_{(1)}\|_*, \alpha_2\|\mathbf{X}_{(2)}\|_*$ and $\alpha_3\|\mathbf{X}_{(3)}\|_*$. The right figure shows the value per iterations of penalty terms for local low-rank tensor, i.e. $\alpha_1 \sum_{\omega\in\Omega} \|\mathcal{R}_\omega(\mathbf{X}_{(1)})\|_*$, $\alpha_2 \sum_{\omega\in\Omega} \|\mathcal{R}_\omega(\mathbf{X}_{(2)})\|_*$ and $\alpha_3 \sum_{\omega\in\Omega} \|\mathcal{R}_\omega(\mathbf{X}_{(3)})\|_*$. Observe the increase of $\alpha_1\|\mathbf{X}_{(1)}\|_*$ and $\alpha_2\|\mathbf{X}_{(2)}\|_*$ for low-rank tensor, while it decreases for local low-rank tensor. This figure illustrates the fact that matrices $\mathbf{X}_{(2)}$ and $\mathbf{X}_{(3)}$ cannot be considered properly as low-rank (while having only a few significant singular values) in contrast to submatrices $\mathcal{R}_\omega(\mathbf{X}_{(2)})$ and $\mathcal{R}_\omega(\mathbf{X}_{(3)})$.

7.4 Discussion

Parallel MRI is implemented on most commercial scanners, and it is the most widely used technique to reduce MRI acquisition time in clinical practice. Thus, it is not only interesting but highly recommended to evaluate the combination of complementary acceleration methods with parallel imaging.

The use of proximal splitting algorithms in this study was particularly helpful because dynamic PMRI generates a much higher volume of data than single coil dynamic MRI. Fast convergent optimisation algorithms lower the number of iterations needed and thus reduce the general computational cost of the algorithm.

A limitation of this work is to only have used a numerical phantom, although this dataset was relatively realistic because of complex-valued noisy data and the presence of motion. Sensitivity maps were assumed to be known within SENSE reconstructions, but in practice only an estimation is possible. In a real scenario where a good estimation of the sensitivity maps may be difficult to obtain for various reasons, it will affect the reconstruction quality of approaches based on sensitivity encoding. Similarly, note that we have used a dataset of dimensions $N_xN_y = 128^2, N_t = 40, N_c = 16$ in these simulations. We have chosen these dimensions to reflect typical values of dynamic PMRI data. However, results may also vary depending on the number of time frames or the number of coils used, especially for the low-rank tensor approach that uses matrices (unfoldings) based on these dimensions.

7.5 Related works

In a nondynamic context, notable techniques that combine PMRI with sparsity prior include CS-SENSE [166] and wavelet-based regularised SENSE [167]. Similarly in a nondynamic context, a method that combines parallel imaging with low-rank structure is SAKE by Shin et al. [168]. This calibrationless method uses the singular value hard thresholding operator $\mathrm{SVT}^{\mathcal{H}}$ on a matrix that is formed by taking the multichannel **k**-space data which proves to have Hankel structure and a low-rank property.

In a dynamic context, apart from the **k**-t SPARSE-SENSE [165], there has been recent works that combine dynamic parallel imaging and both low-rank and sparse constraints such as Refs. [169] and [134]. These works basically extend previous developed methods using low-rank and sparse constraints for single coil dynamic MRI by integrating coil sensitivity information with the standard SENSE model. In Ref. [169], Lingala et al. proposed a direct extension of the **k**-t SLR formulation as in Eq. (3.61) but with coil sensitivities modelling, i.e.

$$\min_{\mathbf{X}} \|\mathcal{E}_{\Gamma}(\mathbf{X}) - \mathbf{y}\|_2^2 + \alpha\|\mathbf{X}\|_{0.1} + \beta\,\mathrm{TV}(\mathbf{X}). \tag{7.13}$$

In Ref. [134], the authors proposed a model that integrates low-rank constraint within the partial separability model (as in PS-Sparse), group sparsity sparse constraint using the $\ell_{1,2}$ norm, and sensitivity encoding.

The tensor framework for dynamic parallel MRI has not received so much attention so far to our knowledge, apart from the original idea proposed by Trzasko and Manduca in Ref. [159].

7.6 Conclusion

In this chapter, we have extended low-rank based recovery methods for dynamic parallel imaging.

We have shown that the low-rank structure could be exploited differently in dynamic multicoil MRI, within low-rank matrix recovery techniques but also considering data within a tensor framework. We have extended and developed fast proximal gradient algorithms to solve various forms of problems combining multicoil dynamic MR imaging and low-rank based constraints.

Chapter 8

Summary and perspectives

Contents

8.1 Contributions

Magnetic resonance imaging is a mature medical imaging technology based on the nuclear magnetic resonance phenomenon that is increasingly used in hospitals to study and interpret the human body, and provide diagnosis of various diseases. During its advancement, MRI has also lead to the development of specific applications such as dynamic, functional or diffusion MRI that enable a more profound exploration and interpretation of the human body. Although tremendous progress has been made since the first MRI pictures were produced in the early 1970s, MRI and its applications still remain limited in many aspects. In particular, the full potential of MRI has not yet been exploited due to fundamental limits in how magnetic resonance imaging collects and reconstructs data.

In this thesis, we have focused on the development of mathematical and computational methods for inverse problems in magnetic resonance imaging, mainly to address the acquisition time constraints of the MRI procedure. This work was primarily inspired by significant developments both from theory and practice in signal recovery techniques from partial data that have emerged in the last past decade in the fields of signal processing and information theory. We have more specifically shown how prior information based on low-dimensional signal models such as low-rank structure and sparseness could help in the reconstruction of spatio-temporal MR images from sub-Nyquist samples arising from a nuclear magnetic resonance experiment. The major contributions of this thesis are as follows.

- The development and characterisation of fast and efficient computational tools

for sub-Nyquist dynamic MRI using convex low-rank and sparse constraints via proximal gradient methods (chapter 5). We have provided important progress in terms of algorithmic developments and performance of these methods through undersampled MRI datasets with relatively realistic characteristics and comparative study of state of the art methods. The proposed proximal gradient framework has proved to provide a flexible computational tool for MRI reconstruction that necessitates to handle nonsmooth penalties in a relatively large-scale setting. We have also shown that while the combination of low-rank structure and sparseness will generally offer superior reconstruction results than using individual constraint, low-rank a priori only is a serious alternative to standard sparse approach for dynamic imaging.

- The development and characterisation of a novel joint reconstruction-separation method via the low-rank plus sparse matrix decomposition technique (chapter 6). This technique goes further than standard reconstruction methods whose principal goals are generally only to recover data the closest possible to the true signal. With the proposed approach, we have departed from such methods to jointly reconstruct and separate data based on their intrinsic characteristics. We have shown this decomposition could be used in a simulated dynamic contrast enhanced sequence where intensity changes over time, to help in motion estimation. While further work is necessary to validate this approach more deeply, this is a first step towards methods that combine directly in the reconstruction process other tasks that can help in the interpretation of images.

- The development and characterisation of rank constraint recovery methods with dynamic parallel MRI (chapter 7). We have extended previous developed method to handle dynamic multicoil imaging and have shown how the low-rank structure could be exploited and combined with dynamic PMRI in different ways.

8.2 Limitations and further directions

There are several limitations of the work presented in this thesis. First, we are guilty of having committed *inverse crimes*, because we have used finite-dimensional signals as references to solve infinite-dimensional inverse problems. Hence, the continuous nature of data has not been taken into account as pointed out in section 3.5. In practice, this omission is largely widespread in the MRI reconstruction literature, and it is surprisingly only recently that researchers have started to look out for solutions of this problem (see for example Ref. [93]). However, since all methods were evaluated within the same framework, these experiments and comparisons are still valuable. In the future though, taking this step into account would better

characterise the behaviour of these algorithms in a more realistic scenario.

Another limitation of this work is that we have mainly used data which were retrospectively undersampled. The reasons we have followed this approach were twofold. First, this allows to have access to a reference signal, the "ground truth", which is extremely helpful to further compare reconstruction algorithms. Second, retrospectively undersampling makes it possible to generate as many sampling trajectories as necessary without the need of requiring a real scanning procedure, which saves a lot of time. Of course, the downside of this approach is that these sampling strategies are only approximate, but these are largely overwhelmed with the aforementioned benefits. Prospectively undersampling is generally more difficult because of limitations on commercial scanners to directly produce undersampled data. However, it gives insights on the true performance of reconstruction algorithms in a realistic procedure and for this reason should not be neglected.

While the image reconstruction literature in MRI has remarkably grown over recent years, there is still an important lack of standard numerical phantoms and common datasets to evaluate MRI reconstruction algorithms from which this thesis may have suffered. Similarly, not all reconstruction algorithms are being made available to the community which complicates reproducible research and generally results in techniques not being evaluated or not being implemented in clinical protocols to be further validated. A positive trend can be observed in the last few years, and notable initiatives have been proposed recently such as

- the design and availability of a realistic cardiac MRI phantom[1] based on XCAT by Wissmann et al. [170],
- the creation of a platform[2] to share fully and undersampled MRI datasets by Lustig and Vasanawala,
- the development of a vendor-independent format data[3] (ISMRMRD) as a prerequisite to share MRI algorithms and codes.

These kind of initiatives can greatly contribute to better shared and reproducible research in the MRI reconstruction community in the future.

In this thesis, we have opted for convex optimisation which has been increasingly used for regularised linear inverse problems. More specifically in signal recovery techniques, convex optimisation is generally employed as a surrogate to solve intractable problems that involve difficult penalties such as the ℓ_0 pseudonorm and/or rank constraints. Convex relaxation hence relies on solving a convex program that only approximates the original problem, although under some assumptions exact signal

[1] `http://www.biomed.ee.ethz.ch/research/bioimaging/cardiac/mrxcat`.
[2] `http://mridata.org`.
[3] `http://ismrmrd.sourceforge.net`.

recovery may actually be possible. There are multiple other approaches that could be examined, first within the proximal splitting framework such as recently developed second-order proximal Newton methods [171] or simply other more standard proximal splitting algorithms as listed in Ref. [100]. Standard convex optimisation alternatives to proximal splitting methods that can handle nonsmooth penalties, such as subgradient and interior point methods, are however limited respectively because of slow global convergence rate and expensive iteration cost as the dimension of the problem increases. More recently, there has been renewed interest in *conditional gradient* method for problems with ℓ_1 and nuclear norms. This method is also known as *Frank-Wolf algorithm* [172] and can be traced back to the mid-1950s. In contrast to first-order methods which solve a projection problem at each iteration, conditional gradient algorithms easily solve a linearisation of the objective function over the feasible set at each iteration. Finally, further investigation would be needed to evaluate greedy approaches for low-rank and sparse reconstruction.

A limitation of iterative optimisation algorithms is the choice of regularisation parameters to provide the adequate trade-off between data fidelity and penalty terms. Since in a practical scenario, the ground truth or reference signal is not available, it will in general only be possible to find a correct regularisation parameter through many trials and visual inspection. Some methods have been developed to overcome this problem such as L-curve or generalised cross-validation methods, but they similarly require to compute multiple solutions of minimisation problems by varying the regularisation parameters. Automatic selection of the right regularisation parameter has been proposed recently based on Stein's unbiased risk estimate, see Refs. [141, 145, 173], but it is limited to specific inverse problems and therefore remains an open issue in the general case.

Methods developed in this thesis were based on regularisation prior promoting low-complexity models. The main justification is that very often, the underlying information of high-dimensional data can actually lie in a much lower dimensional space. We have specifically investigated sparseness and low-rankdness (matrix and tensors), but there are other potential low-dimensional signal models that could be examined. In particular, methods and algorithms developed in this thesis were training-free, but important advancements could be made by learning low-dimensional signal models, although as of today these methods are still computationally expensive. We should also expect to see more work that could be referred to as *intelligent reconstruction* that goes in the line of the joint reconstruction-separation approach of chapter 6, where advanced interpretation tasks such as segmentation, classification, or registration are more tightly connected together within the reconstruction process itself.

In this thesis, we have also combined complementary acceleration approaches. We have mainly shown how we could combine parallel MRI technique that uses

coil sensitivity informations with a low-dimensional signal model based on low-rank structure. A direction for future work would be to study the influence of both low-rank and sparse constraints within the parallel MRI framework, whether using the SENSE model or the low-rank tensor framework. Recently, Goldfarb and Qin [174] have proposed the *robust* low-rank tensor recovery, an extension of RPCA as in Eq. (6.2) within the tensor framework. The tensor to recover is assumed to be the sum of a low-rank tensor plus a sparse one, i.e. $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} + \boldsymbol{\mathcal{S}}$. This could be another interesting direction for future work, where the convex higher-order RPCA problem for undersampled multichannel dynamic MRI could be written as

$$\min_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}} \Big\{ \sum_{q=1}^{Q} \alpha_q \sum_{\omega\in\Omega} \|\mathcal{R}_\omega(\mathbf{L}_{(q)})\|_* + \|\boldsymbol{\mathcal{S}}\|_1 \Big\} \quad \text{s.t.} \quad \mathcal{E}(\boldsymbol{\mathcal{L}} + \boldsymbol{\mathcal{S}}) = \mathbf{y}, \tag{8.1}$$

where the low-rank part is promoted locally. This model may prove useful for example to isolate remaining noise artefacts in the sparse tensor component and hence potentially obtain higher reconstruction results through better denoising.

Although the design and inclusion of regularisation priors that better reflect the data are desirable, note that a priori information about the structure only is generally not sufficient to transform the problem into a well-posed one. Sparse and low-rank based signal recovery guarantees also rely on specific sensing procedures with notions such as low coherence or restricted isometry property. In this thesis, these concepts were empirically achieved using an MRI encoding operator that was designed to sample randomly the Fourier space in accordance with the energy distribution. A limitation of this thesis is that we have not much focused our attention on the relation between sampling strategies and theoretical recovery guarantees. We have chosen heuristic $(\mathbf{k}, t)$-space sampling trajectories that prove to be both feasible from an MRI acquisition scenario and satisfying regarding empirical reconstruction results from partial data, instead of nonrealistic approaches that maximise the probability of compressed sensing recovery guarantees.

Most important MRI problems are not specific to MRI itself or even medical imaging. For example, MR image analysis and reconstruction share fundamentally similar mathematical open problems that can be found in many other areas. Therefore, research in magnetic resonance imaging enjoys multidisciplinary contributions from various communities such as medicine, physics, computer science or electrical engineering. Continuous developments from multiple fields are expected to constantly improve MR imaging systems, producing images that will maximise the quantity, accuracy and precision of information, thus enabling superior understanding of biological systems and extending human knowledge.

Appendix A

Mathematical background

This appendix briefly reviews some relevant mathematical concepts. We refer the reader to Refs. [95, 96, 175] for more details.

A.1 Vector spaces

In this dissertation, functions that represent a mathematical description of a physical process (i.e. signals) can be regarded as vectors, and as such they lie in a *vector space*. Formally, a vector space is a set of elements called vectors together with two operations, addition and scalar multiplication, that are assumed to satisfy some properties (axioms) such as commutative, associative or distributive laws.

Typical examples of vector spaces used in this thesis are the N-dimensional Euclidean space $\mathbb{R}^N$ and the space of complex-valued N-dimensional vectors $\mathbb{C}^N$. Finite-dimensional and discrete signals can be though as vectors in $\mathbb{R}^N$ or $\mathbb{C}^N$ composed of the samples of the (continuous) function they represent.

Two important geometric notions can be defined on vector spaces: *norm* and *inner product*. A *normed vector space* is a vector space on which there is defined a real-valued function called the norm which maps each elements of the vector space into a real number called the norm. The norm must satisfy some specific axioms (positive definiteness, homogeneity and triangle inequality) and can be interpreted as an abstraction of the concept of length. By introducing this measure of distance, topological concepts such as openness, closure or convergence can be defined. A vector space equipped with an inner product is called an *inner product space*. Since an inner product naturally induces a norm, inner product spaces are also normed vector spaces. The inner product on a vector space is a real or complex-valued function that must satisfy some axioms (distributivity, linearity in the first argument, Hermitian symmetry, positive definiteness). Inner products are particularly useful to define the notion of orthogonality for example.

The standard norm on $\mathbb{C}^N$ is the Euclidean (or ℓ_2) norm. Consider $\mathbf{x} \in \mathbb{C}^N$, the

ℓ_2 norm of $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{n=1}^{N} |x_n|^2}, \tag{A.1}$$

and the standard inner product on $\mathbb{C}^N$ between $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathrm{H}} \mathbf{y} = \sum_{n=1}^{N} x_n y_n^*. \tag{A.2}$$

More generally, we can define the ℓ_p norm as

$$\|\mathbf{x}\|_p = \left(\sum_{n=1}^{N} |x_n|^p \right)^{1/p} \tag{A.3}$$

for $p \in [1, \infty)$. For $p = 1$, this yield the ℓ_1 norm

$$\|\mathbf{x}\|_1 = \sum_{n=1}^{N} |x_n|, \tag{A.4}$$

and for $p = \infty$, the ℓ_∞ norm is defined as

$$\|\mathbf{x}\|_\infty = \max_{n=1\ldots N} |x_n|. \tag{A.5}$$

When $0 < p < 1$, these functions are called *quasinorms* because positive definiteness and homogeneity properties hold, but the triangle inequality is replaced by a (weaker) quasitriangle inequality. When $p = 0$, it is neither a norm nor a quasinorm, but it is sometimes referred to as a *pseudonorm*. It is defined as

$$\|\mathbf{x}\|_0 = \#\{n : x_n \neq 0\}, \tag{A.6}$$

which is the number of nonzero elements in $\mathbf{x}$. Consider the unit ball of the ℓ_p function defined by

$$\mathcal{B}_p = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_p \leq 1\}. \tag{A.7}$$

Figure A.1 illustrates the shape of such balls for various p in $\mathbb{R}^2$ and shows that norms are convex functions while quasinorms are not (reflecting the violation of the triangle inequality). This figure also shows that the ℓ_1 norm is the closest convex norm to the ℓ_0 pseudonorm.

For $M \times N$ matrices, the Schatten p-norm is defined as

$$\|\mathbf{X}\|_p = \left(\sum_{n=1}^{\min(M,N)} \sigma_n^p \right)^{1/p} \tag{A.8}$$

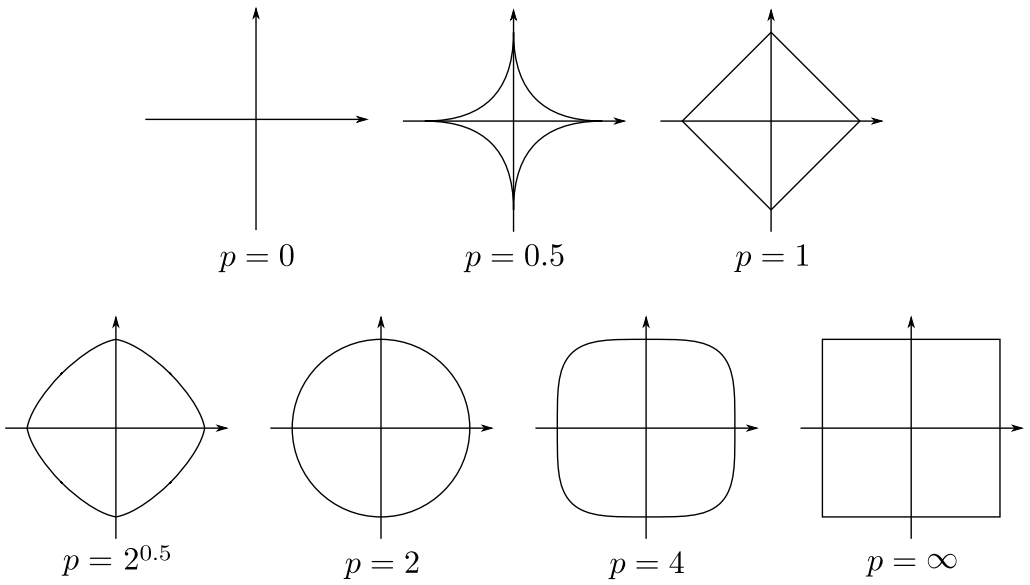

Figure A.1: Unit balls for ℓ_p functions in $\mathbb{R}^2$ for various p.

for $p \in [1, \infty)$, where σ_n denotes the n^{th} singular value of $\mathbf{X}$. Important cases are $p = 2$ which yields the Frobenius norm,

$$\|\mathbf{X}\|_F = \sqrt{\sum_{n=1}^{\min(M,N)} \sigma_n^2} = \sqrt{\sum_{m=1}^{M} \sum_{n=1}^{N} |x_{mn}|^2}, \tag{A.9}$$

and $p = 1$ which defines the nuclear norm (also known as trace norm),

$$\|\mathbf{X}\|_* = \sum_{n}^{\min(M,N)} \sigma_n. \tag{A.10}$$

The nuclear norm can be regarded as the ℓ_1 norm of the vector of singular values, and it is the convex envelope of the matrix rank [42, 45].

We conclude this brief review with the notion of *Hilbert* and *Banach spaces*. A normed vector space that is *complete* is called a Banach space, and a normed vector space with an inner product that is *complete* is called an Hilbert space. In short, completeness means that every Cauchy sequence in the vector space converges to a vector. For example, it can be shown that the space of complex-valued N-dimensional vectors $\mathbb{C}^N$ is a complete vector space. Equipped with the ℓ_2 norm and an inner product, such a vector space in an Hilbert space.

A.2 Convex optimisation

Although we use real vector spaces in the following definitions, sets in and functions on $\mathbb{C}^N$ can be identified with $\mathbb{R}^{2N}$.

A.2.1 Convex sets

A subset $C \subseteq \mathbb{R}^N$ is said to be *convex* if we have

$$\forall \mathbf{x}, \mathbf{y} \in C, \forall \theta \in [0,1], \quad \theta\mathbf{x} + (1-\theta)\mathbf{y} \in C. \tag{A.11}$$

This means that given two points in C, the line segment between them lies in C.

A.2.2 Convex functions

The *domain* of a function $f : \mathbb{R}^N \to \mathbb{R}$ is the set $\operatorname{dom} f \subseteq \mathbb{R}^N$ over which f is well-defined, i.e.

$$\operatorname{dom} f \equiv \{\mathbf{x} \in \mathbb{R}^N : -\infty < f(\mathbf{x}) < +\infty\}. \tag{A.12}$$

A function $f : \mathbb{R}^N \to \mathbb{R}$ is convex if $\operatorname{dom} f$ is a convex set and if

$$\forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} f, \forall \theta \in [0,1], \quad f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}). \tag{A.13}$$

The function f is called *stricly* convex if

$$\forall \mathbf{x} \neq \mathbf{y}, \forall \theta \in (0,1), \quad f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) < \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}). \tag{A.14}$$

The convexity of the function can also be defined using the epigraph, which is intuitively the domain of the function "above the graph": the function f is convex if and only if its epigraph is a convex set. The epigraph makes the connection between convex sets and convex functions. Alternative ways to define convexity of a function is to use first-order (gradient) or second-order condition (Hessian). If f is differentiable, the function f is convex if

$$\forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} f, \quad f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}), \tag{A.15}$$

and strictly convex if

$$\forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} f, \mathbf{x} \neq \mathbf{y}, \quad f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}). \tag{A.16}$$

If f is twice differentiable, then the function f is convex if and only if its Hessian $\nabla^2 f(\mathbf{x})$ is positive semidefinite everywhere on $\operatorname{dom} f$, i.e.

$$\forall \mathbf{x} \in \operatorname{dom} f, \quad \nabla^2 f(\mathbf{x}) \succeq 0, \tag{A.17}$$

where $\succeq$ denotes the componentwise inequality. (For strict convexity, $\succeq$ is replaced by $\succ$.)

An intuitive way to understand convexity is graphically by considering a function of one variable $f : \mathbb{R} \to \mathbb{R}$, as illustrated in figure A.2. The function f is convex if

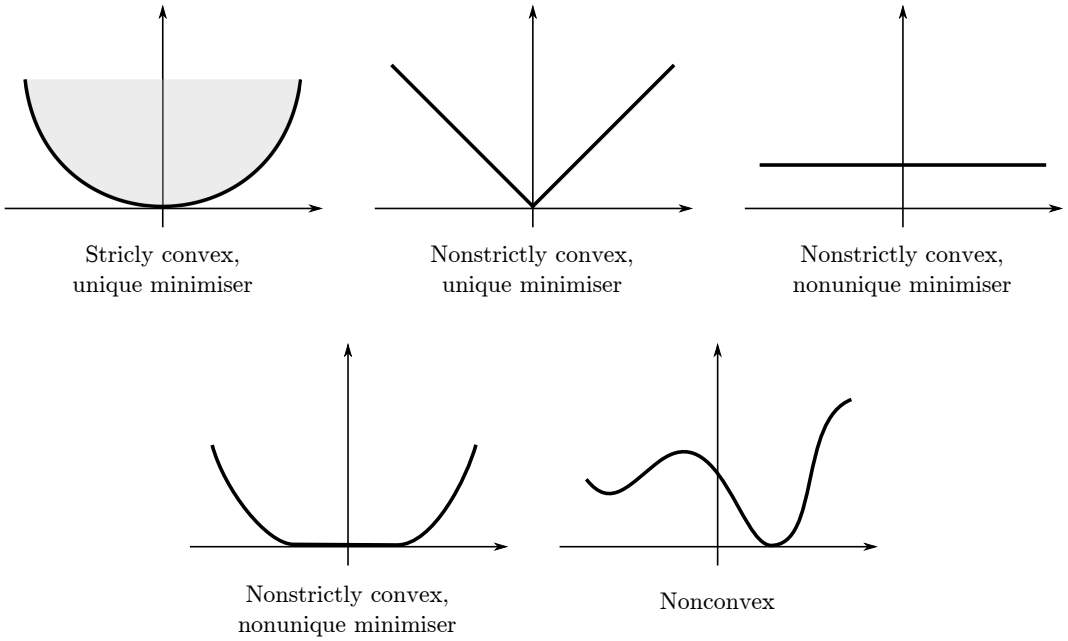

Figure A.2: Illustration of convex and nonconvex real-valued functions of one variable. For the strictly convex function (top left), the epigraph is shown in gray.

the line segment connecting any two points on the graph of f lies above the graph. For strict convexity, the line segment connecting any two points on the graph of f lies strictly above the graph (i.e. there is always a gap between the function graph and the line segment connecting two points on the graph).

A.2.3 Convex optimisation problems

Formally, a convex optimisation problem is one of the form

$$\min_{\mathbf{x}} f_0(\mathbf{x}) \quad \text{s.t.} \quad \begin{cases} f_j(\mathbf{x}) \leq 0, & j = 1, \ldots, J, \\ h_k(\mathbf{x}) = 0, & k = 1, \ldots, K, \end{cases} \tag{A.18}$$

where $\mathbf{x} \in \mathbb{R}^N$ is the optimisation variable, $f_0 : \mathbb{R}^N \to \mathbb{R}$ is the convex objective function, $f_j : \mathbb{R}^N \to \mathbb{R}$ are convex inequality constraint functions, and $h_k : \mathbb{R}^N \to \mathbb{R}$ are affine equality constraint functions.

A point $\mathbf{x}$ is said *feasible* if it satisfies the constraints. A feasible point for which the minimum is attained is called a *minimiser* or an *optimal point*. It is usually denoted by $\mathbf{x}^\star$, and thus $f_0(\mathbf{x}^\star)$ is the *optimal value*. For convex problems, any locally optimal point is globally optimal. Note that to have a *unique* minimiser, a convex optimisation problem should in general be strictly convex. A simple example of a convex function having multiple minimisers is a constant function as shown in figure A.2.

Convex optimisation problems include linear programming problems of the form

$$\min_{\mathbf{x}} f_0(\mathbf{x}) \quad \text{s.t.} \quad f_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, J, \tag{A.19}$$

where the objective function f_0 and the inequality constraints functions f_j are all affine. Convex optimisation problems also include quadratic programming problems of the form

$$\min_{\mathbf{x}} f_0(\mathbf{x}) \equiv \mathbf{c}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{Q}\mathbf{x} \quad \text{s.t.} \quad f_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, J, \tag{A.20}$$

for $\mathbf{c} \in \mathbb{R}^N$, where $\mathbf{Q}$ is a positive semidefinite matrix (to ensure the function f_0 is convex) and f_j are affine inequality constraints functions.

A.2.4 Gradient method

To minimise a convex differentiable function, one of the most simple algorithm is the *gradient method*. It consists in trying to decrease the value of the function by taking a step in the direction of the negative gradient, i.e.

$$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^k - \rho^k \nabla f(\mathbf{x}^k) \tag{A.21}$$

where ρ^k is the step size and ∇f is the gradient of f.

Gradient methods suffer from slow convergence, so more sophisticated techniques have been developed such as quasi-Newton, conjugate gradient and accelerated gradient methods. Similarly, since classical gradient methods cannot handle nonsmooth problems, techniques have been developed such as subgradient, smoothing or proximal gradient methods. A notable use of the smoothing technique to handle the nondifferentiability of the ℓ_1 norm is to replace it by the Huber function, a differentiable function that is close to the absolute value function.

A.3 Differentiability, smoothness and Lipschitz continuity

We recall that a function f is said to be of class C^k if the derivatives up to the k^{th} order exist and are continuous. The class C^0 denotes the space of continuous functions, and the class C^1 the space of continuously differentiable functions. The class C^∞ refers to functions that have derivatives of all orders.

A function f is generally referred to as *smooth* when it is differentiable (up to some desired order) with continuous derivatives. *Nonsmooth* optimisation problems for example refer to problems that include at least one nondifferentiable function.

When a (differentiable) function $f : \mathbb{C}^N \to \mathbb{R}$ is assumed to belong to the class of

Lipschitz continuous gradient functions, it means that the gradient ∇f is Lipschitz continuous with some constant $L > 0$, i.e. it satisfies

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{z})\|_2 \leq L\|\mathbf{x} - \mathbf{z}\|_2, \quad \forall \mathbf{x}, \mathbf{z} \in \mathbb{C}^N. \tag{A.22}$$

This class of functions is often denoted $C_L^{1,1}$ where the second "1" in the superscript refers to the order of the derivative that is L-Lipschitz continuous. When f is twice differentiable, an equivalent of this definition is that the largest eigenvalue of the Hessian $\nabla^2 f(\mathbf{x})$ is upper bounded by L.

For the linear and Gaussian case as in MRI, the data fidelity term is $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2$, whose gradient $\nabla f(\mathbf{x}) = \mathbf{E}^{\mathrm{H}}(\mathbf{E}\mathbf{x} - \mathbf{y})$ satisfies

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{z})\|_2 = \|\mathbf{E}^{\mathrm{H}}\mathbf{E}(\mathbf{x} - \mathbf{z}))\|_2 \leq \|\mathbf{E}\|_2^2\|\mathbf{x} - \mathbf{z}\|_2, \quad \forall \mathbf{x}, \mathbf{z} \in \mathbb{C}^N, \tag{A.23}$$

so that ∇f is Lipschitz continuous with constant

$$L \leq \|\mathbf{E}\|_2^2, \tag{A.24}$$

where $\|\mathbf{E}\|_2$ denotes the spectral norm of $\mathbf{E}$. The spectral norm is equal to the largest singular value of $\mathbf{E}$ or equivalently the square root of the largest eigenvalue $\lambda_{\max}$ of the positive-semidefinite matrix $\mathbf{E}^{\mathrm{H}}\mathbf{E}$, i.e.

$$\|\mathbf{E}\|_2 = \sigma_{\max}(\mathbf{E}) = \sqrt{\lambda_{\max}(\mathbf{E}^{\mathrm{H}}\mathbf{E})}, \tag{A.25}$$

assuming σ_i and λ_i represents respectively the singular values and eigenvalues. The largest singular value of a normalised orthogonal discrete Fourier matrix is one, which means that the Lipschitz constant should be chosen as

$$L \leq 1 \tag{A.26}$$

in our MRI reconstruction problems. The Lipschitz constant is useful because when known, the step size in proximal optimisation algorithms can be easily determined to achieve convergence, although strategies exist when it is not computable.

Appendix B

Error metrics

In this thesis, performance of reconstruction methods are often quantified using the following expression

$$-10\log_{10}\Big(\frac{\|\widehat{\mathbf{x}}-\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}\Big), \tag{B.1}$$

where $\mathbf{x}$ (resp. $\widehat{\mathbf{x}}$) represents the reference noiseless signal (resp. estimated signal). When considering matrices, the Frobenius norm instead of the ℓ_2 norm can be used equivalently,

$$-10\log_{10}\Big(\frac{\|\widehat{\mathbf{X}}-\mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2}\Big), \tag{B.2}$$

where $\mathbf{X}$ (resp. $\widehat{\mathbf{X}}$) represents the ground truth noiseless matrix (resp. estimated matrix). These quantities are expressed in decibels for convenience.

In the case of dynamic imaging, the above expressions consider both space and time informations. An alternative error metric is the *normalised mean square error* (NMSE) computed at each time frame n,

$$\frac{\|\widehat{\mathbf{x}}_n-\mathbf{x}_n\|_2^2}{\|\mathbf{x}_n\|_2^2}, \tag{B.3}$$

where $\mathbf{x}_n$ (resp. $\widehat{\mathbf{x}}_n$) represents the reference noiseless image (resp. estimated image) at time frame n.

Bibliography

[1] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: Physical principles and sequence design.* Wiley-Liss, 1999. 33, 46

[2] Z.-P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging: A signal processing perspective.* IEEE Press, 2000. 33, 41, 42

[3] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince. *MRI: From picture to proton.* Cambridge University Press, 2006. 33

[4] P. C. Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973. 39

[5] S. Ljunggren. A simple graphical representation of Fourier-based imaging methods. *Journal of Magnetic Resonance*, 54(2):338–343, September 1983. 40

[6] D. B. Twieg. The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical Physics*, 10(5), 1983. 40

[7] J. B. Weaver, Y. Xu, D. M. Healy, and J. R. Driscoll. Wavelet-encoded MR imaging. *Magnetic Resonance in Medicine*, 24:275–87, April 1992. 41

[8] D. M. Healy and J. B. Weaver. Two applications of wavelet transforms in magnetic resonance imaging. *IEEE Transactions on Information Theory*, 38(2):840–860, 1992. 41

[9] G. P. Zientara, L. P. Panych, and F. A. Jolesz. Dynamically adaptive MRI with encoding by singular value decomposition. *Magnetic Resonance in Medicine*, 32(2):268–74, August 1994. 41

[10] Y. Cao and D. N. Levin. On the relationship between feature-recognizing MRI and MRI encoded by singular value decomposition. *Magnetic Resonance in Medicine*, 33(1):140–2, January 1995. 41

[11] J. A. Fessler. Reconstruction from Fourier samples (gridding and alternatives). In *Image reconstruction: Algorithms and analysis.* Unpublished book draft. 41

[12] C. L. Epstein. *Introduction to the mathematics of medical imaging.* SIAM, 2008. 42

[13] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 42

[14] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. 42

[15] H. D. Lüeke. The origins of the sampling theorem. *IEEE Communications Magazine*, 37(4):106–108, 1999. 42

[16] J. M. Pauly. Gridding & the NUFFT for non-Cartesian image reconstruction. In *ISMRM Educational Course on Image Reconstruction*, 2012. 45

[17] A. Dutt and V. Rokhlin. Fast Fourier transforms for nonequispaced data. *SIAM Journal on Scientific computing*, 14(6):1368–1393, 1993. 45

[18] J. A. Fessler and B. P. Sutton. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, 2003. 45

[19] J. A. Fessler. On NUFFT-based gridding for non-Cartesian MRI. *Journal of Magnetic Resonance*, 188(2):191–5, October 2007. 45

[20] J. B. Johnson. Thermal agitation of electricity in conductors. *Physical Review*, 32:97–109, 1928. 46

[21] H. Nyquist. Thermal agitation of electric charge in conductors. *Physical Review*, 32:110–113, 1928. 46

[22] G. A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Magazine*, 14(1):56–66, 1997. 46

[23] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914, 2005. 47

[24] Q.-S. Xiang and R. M. Henkelman. K-space description for MR imaging of dynamic objects. *Magnetic Resonance in Medicine*, 29(3):422–428, 1993. 48

[25] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging.* Institute of Physics Publishing, 1998. 51

[26] C. R. Vogel. *Computational methods for inverse problems.* Society for Industrial and Applied Mathematics, 2002. 51

[27] P. C. Hansen. *Discrete inverse problems: Insight and algorithms.* Society for Industrial and Applied Mathematics, 2010. 51, 68

[28] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, XIII(4):49–52, 1902. 51

[29] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006. 53

[30] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 53

[31] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, December 2005. 55, 58

[32] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006. 55

[33] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–95, December 2007. 56, 63, 95

[34] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, pages 72–82, 2008. 56, 94

[35] H. L. Taylor, S. C. Banks, and J. F. McCoy. Deconvolution with the l1 norm. *Geophysics*, 44(1):39–52, 1979. 56

[36] S. Levy and P. K. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46(9):1235–1243, 1981. 56

[37] C. Walker and T. J. Ulrych. Autoregressive recovery of the acoustic impedance. *Geophysics*, 48(10):1338–1350, 1983. 56

[38] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996. 56

[39] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 1998. 56

[40] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 57

[41] J. J. Gerbrands. On the relationships between SVD, KLT and PCA. *Pattern Recognition*, 14:375–381, 1981. 57

[42] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002. 57, 167

[43] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011. 58

[44] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211–231, 2009. 58

[45] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010. 58, 59, 167

[46] B. Recht, W. Xu, and B. Hassibi. Null space conditions and thresholds for rank minimization. *Mathematical Programming*, 127(1):175–202, 2011. 58

[47] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9-10):589–592, May 2008. 58

[48] T. T. Cai and A. Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 35(1):74–93, July 2013. 58, 59

[49] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013. 59, 74

[50] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, December 2006. 59

[51] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constructive Approximation*, 28(3):277–289, February 2008. 59

[52] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, January 2008. 59

[53] Z. Lin, M. Chen, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report, University of Illinois at Urbana-Champaign, 2009. 59, 122

[54] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. Technical report, University of Illinois at Urbana-Champaign, 2009. 59

[55] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21, 2011. 59, 120

[56] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), 2011. 59, 60, 120, 122

[57] X. M. Yuan and J. F. Yang. Sparse and low-rank matrix decomposition via alternating direction methods. *Pacific Journal of Optimization*, 9(1):167–180, 2013. 59, 120, 122

[58] M. Tao and X. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21(1):57–81, 2011. 60, 120

[59] A. E. Waters, A. C. Sankaranarayanan, and R. G. Baraniuk. SpaRCS: Recovering low-rank and sparse matrices from compressive measurements. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 1089–1097, Granada, 2011. 60, 87, 120

[60] X. Li. Compressed sensing and matrix completion with constant proportion of corruptions. *Constructive Approximation*, 37(1):73–99, December 2012. 60, 120

[61] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis. Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory*, 59(7):4324–4337, July 2013. 60, 120

[62] J. Wright, A. Ganesh, K. Min, and Y. Ma. Compressive principal component pursuit. *Information and Inference*, 2(1):32–68, June 2013. 60, 120

[63] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455, 2009. 60, 61

[64] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, January 2012. 61

[65] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. *arXiv*, 2011. 61

[66] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, February 2011. 61, 150

[67] P. Mansfield. Multiplanar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55–L58, 1977. 62

[68] A. Haase, J. Frahm, D. Matthaei, W. Hänicke, and K.-D. Merboldt. FLASH imaging: Rapid NMR imaging using low flip-angle pulses. *Journal of Magnetic Resonance*, 67:258–266, December 1986. 62

[69] J. Tsao and S. Kozerke. MRI temporal acceleration techniques. *Journal of Magnetic Resonance Imaging*, 36(3):543–560, September 2012. 62, 65, 128

[70] B. Madore, G. H. Glover, and N. J. Pelc. Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. *Magnetic Resonance in Medicine*, 42(5):813–28, November 1999. 62, 67

[71] J. Tsao, P. Boesiger, and K. P. Pruessmann. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine*, 50(5):1031–42, November 2003. 63, 67, 148, 153

[72] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly. k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, page 2003, 2006. 63, 67, 114, 123

[73] H. Jung, J. C. Ye, and E. Y. Kim. Improved k-t BLAST and k-t SENSE using FOCUSS. *Physics in Medicine and Biology*, 52(11):3201–26, June 2007. 63, 67, 90, 100, 103, 123, 130

[74] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye. k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. *Magnetic Resonance in Medicine*, 61(1):103–116, January 2009. 63, 90, 100, 103, 130

[75] H. Jung, J. Park, J. Yoo, and J. C. Ye. Radial k-t FOCUSS for high-resolution cardiac cine MRI. *Magnetic Resonance in Medicine*, 63(1):68–78, January 2010. 63, 90, 100, 103, 130

[76] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616, March 1997. 63

[77] M. Usman, C. Prieto, T. Schaeffter, and P. G. Batchelor. k-t group sparse: a method for accelerating dynamic MRI. *Magnetic Resonance in Medicine*, 66(4):1163–76, October 2011. 65

[78] Z.-P. Liang. Spatiotemporal imaging with partially separable functions. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 988–991, Washington, 2007. 65

[79] J. P. Haldar and Z.-P. Liang. Spatiotemporal imaging with partially separable functions: A matrix recovery approach. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 716–719, Rotterdam, 2010. 65

[80] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang. Low rank matrix recovery for real-time cardiac MRI. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 996–999, Rotterdam, 2010. 65

[81] J. P. Haldar and D. Hernando. Rank-constrained solutions to linear matrix equations using PowerFactorization. *IEEE Signal Processing Letters*, 16(7):584–587, January 2009. 66

[82] S. G. Lingala, Y. Hu, E. V. R. DiBella, and M. Jacob. Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. *IEEE Transactions on Medical Imaging*, 30(5):1042–1054, May 2011. 66, 90, 91, 97, 100, 103, 130

[83] Y. Hu, S. G. Lingala, and M. Jacob. A fast majorize-minimize algorithm for the recovery of sparse and low-rank matrices. *IEEE Transactions on Image Processing*, 21(2):742–53, February 2012. 66, 90

[84] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 262–265, Boston, June 2009. 66, 87, 111

[85] R. Chartrand. Nonconvex splitting for regularized low-rank + sparse decomposition. *IEEE Transactions on Signal Processing*, 60(11):5810–5819, 2012. 66, 87, 111

[86] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang. Image reconstruction from highly undersampled (k,t)-space data with joint partial separability and sparsity constraints. *IEEE Transactions on Medical Imaging*, 31(9):1809–20, September 2012. 66, 90, 91, 100, 103

[87] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946, 1995. 67

[88] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008. 67

[89] R. Otazo, E. Candès, and D. K. Sodickson. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magnetic Resonance in Medicine*, 2014. 67, 90, 142

[90] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge. Dynamic MR image reconstruction–separation from undersampled (k,t)-space via low-rank plus sparse prior. *IEEE Transactions on Medical Imaging*, 33(8):1689–1701, 2014. 67, 120

[91] J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504, January 2007. 68

[92] B. Adcock, A. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. *arXiv*, 2013. 68, 95

[93] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 31(3):626–636, 2012. 69, 160

[94] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I*, volume 305. Springer, 1993. 71

[95] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. 71, 165

[96] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer, 2004. 71, 165

[97] V. Chandrasekaran, B. Recht, P. A Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, October 2012. 72

[98] D. P. Palomar and Y. C. Eldar. *Convex optimization in signal processing and communications*. Cambridge University Press, 2010. 72

[99] S. Sra, S. Nowozin, and S. J. Wright. *Optimization for machine learning*. MIT Press, 2012. 72

[100] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011. 72, 77, 115, 116, 162

[101] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014. 72

[102] Y. Nesterov and A. Nemirovski. On first-order algorithms for l1/nuclear norm minimization. *Acta Numerica*, 22:509–575, April 2013. 72

[103] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 72, 79

[104] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1 - Theory. *IEEE Transactions on Medical Imaging*, 1(2):81–94, January 1982. 73

[105] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes rendus de l'Académie des Sciences*, 255:2897–2899, 1962. 73

[106] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965. 73

[107] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20:1956–1982, 2010. 76

[108] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005. 76

[109] Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, Université catholique de Louvain, 2007. 77, 90, 115, 116

[110] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, January 2009. 77, 90, 93, 115, 116

[111] A. Beck and M. Teboulle. Gradient-based algorithms with applications to signal-recovery problems. In *Convex optimization in signal processing and communications*. Cambridge University Press, 2010. 77, 78

[112] L. Vandenberghe. Lecture notes EE236C "Optimization methods for large-scale systems", University of California, Los Angeles. 2013–2014. 77

[113] D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. 78

[114] J. Huang, S. Zhang, H. Li, and D. Metaxas. Composite splitting algorithms for convex optimization. *Computer Vision and Image Understanding*, 115(12):1610–1622, December 2011. 78, 150

[115] H. Raguet, J. Fadili, and G. Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013. 78

[116] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976. 79

[117] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975. 79

[118] T. Goldstein and S. Osher. The split Bregman method for L1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009. 79

[119] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010. 79

[120] J. Yang and Y. Zhang. Alternating direction algorithms for l1-problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, 2011. 79

[121] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–95, March 2011. 79

[122] S. Ramani and J. A. Fessler. Parallel MR image reconstruction using augmented Lagrangian methods. *IEEE Transactions on Medical Imaging*, 30(3):694–706, 2011. 79

[123] J. E. Esser. *Primal dual algorithms for convex models and applications to image restoration, registration and nonlocal inpainting*. PhD thesis, University of California, Los Angeles, 2010. 79

[124] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10):707–710, 2007. 87

[125] J. Trzasko and A. Manduca. Highly undersampled magnetic resonance image reconstruction via homotopic l0-minimization. *IEEE Transactions on Medical Imaging*, 28(1):106–121, 2009. 87

[126] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007. 87

[127] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, November 2009. 87

[128] T. Blumensath and M.E. Davies. Normalized iterative hard thresholding: Guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309, April 2010. 87

[129] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, May 2009. 87

[130] P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection. In *Advances in Neural Information Processing Systems (NIPS)*, pages 937–945, 2010. 87

[131] K. Lee and Y. Bresler. ADMiRA: Atomic decomposition for minimum rank approximation. *IEEE Transactions on Information Theory*, 56(9):4402–4416, 2010. 87

[132] J. Tanner and K. Wei. Normalized iterative hard thresholding for matrix completion. *SIAM Journal on Scientific Computing*, 35(5):104–125, 2013. 87

[133] H. Gao, Y. Lin, C. B. Ahn, and O. Nalcioglu. PRISM: A divide-and-conquer low-rank and sparse decomposition model for dynamic MRI. Technical report, University of California, Los Angeles, 2011. 90, 142

[134] A. G. Christodoulou, H. Zhang, B. Zhao, T. Kevin Hitchens, C. Ho, and Z.-P. Liang. High-resolution cardiovascular MRI by integrating parallel imaging with low-rank and sparse modeling. *IEEE Transactions on Biomedical Engineering*, 60(11):3083–3092, 2013. 90, 157

[135] B. Trémoulhéac, D. Atkinson, and S. R. Arridge. Fast dynamic MRI via nuclear norm minimization and accelerated proximal gradient. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 322–325, San Francisco, 2013. 91, 106

[136] B. Trémoulhéac, D. Atkinson, and S. R. Arridge. Low-rank and (x-f)-space sparsity via fast composite splitting for accelerated dynamic MR imaging. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 649–652, Beijing, 2014. 91

[137] K. C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(3):615–640, 2010. 93, 116

[138] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel. An optimal radial profile order based on the Golden Ratio for time-resolved MRI. *IEEE Transactions on Medical Imaging*, 26(1):68–76, January 2007. 96

[139] R. W. Chan, E. A. Ramsay, E. Y. Cheung, and D. B. Plewes. The influence of radial undersampling schemes on compressed sensing reconstruction in breast MRI. *Magnetic Resonance in Medicine*, 67(2):363–77, February 2012. 96

[140] B. Sharif and Y. Bresler. Physiologically improved NCAT phantom (PINCAT) enables in-silico study of the effects of beat-to-beat variability on cardiac MR. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, page 3418, 2007. 97

[141] E. J. Candès, C. A Sing-Long, and J. D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on Signal Processing*, 61(19):4643–4657, 2013. 97, 116, 162

[142] J. Trzasko and A. Manduca. Local versus global low-rank promotion in dynamic MRI series reconstruction. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, page 4371, 2011. 113, 115, 154

[143] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–41, May 2011. 115

[144] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal. Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Transactions on Medical Imaging*, 33(4):979–94, 2014. 115

[145] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. A. Fessler. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods. *IEEE Transactions on Image Processing*, 21(8):3659–72, August 2012. 116, 162

[146] B. Trémoulhéac, D. Atkinson, and S. R. Arridge. Motion and contrast enhancement separation model reconstruction from partial measurements in dynamic

MRI. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on Sparsity Techniques in Medical Imaging*, Nice, 2012. 120

[147] P. S. Tofts. T1-weighted DCE imaging concepts: Modelling, acquisition and analysis. *MAGNETOM Flash (Siemens)*, (3):30–39, 2010. 133, 136

[148] V. Hamy, N. Dikaios, S. Punwani, A. Melbourne, A. Latifoltojar, J. Makanyanga, M. Chouhan, E. Helbren, A. Menys, S. Taylor, and D. Atkinson. Respiratory motion correction in dynamic MRI using robust data decomposition registration - Application to DCE-MRI. *Medical Image Analysis*, 18(2):301–313, 2014. 138

[149] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98(3):278–84, June 2010. 138

[150] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–21, August 1999. 138

[151] W. C. Karl. Regularization in image restoration and reconstruction. In A. C. Bovik, editor, *Handbook of Image and Video Processing*, chapter 3.6, pages 183–202. Elsevier, 2005. 140

[152] A. E. Campbell-Washburn, D. Atkinson, O. Josephs, M. F. Lythgoe, R. J. Ordidge, and D. L. Thomas. Correction of RF spike noise in MR images using Robust Principal Component Analysis. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, page 3780, Salt Lake City, 2013. 141

[153] B. Trémoulhéac, P. G. Batchelor, A. Menys, and D. Atkinson. Decomposition of dynamic MR images with low-rank and sparse matrix separation. In *Proceedings of British Chapter of International Society for Magnetic Resonance in Medicine (BC-ISMRM)*, page 27, Manchester, 2011. 142

[154] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao. Robust principal component analysis-based four-dimensional computed tomography. *Physics in Medicine and Biology*, 56(11):3181–98, June 2011. 142

[155] H. Gao, S. Rapacchi, D. Wang, J. Moriarty, C. Meehan, J. Sayre, G. Laub, P. Finn, and P. Hu. Compressed sensing using prior rank, intensity and sparsity model (PRISM): Applications in cardiac cine MRI. In *Proceedings of*

International Society for Magnetic Resonance in Medicine (ISMRM), page 2242, Melbourne, 2012. 142

[156] H. Gao. Prior rank, intensity and sparsity model (PRISM): a divide-and-conquer matrix decomposition model with low-rank coherence and sparse variation. In *Proceeding of SPIE 8506, Developments in X-Ray Tomography VIII, 85060Y*, October 2012. 142

[157] H. Gao, L. Li, and X. Hu. Compressive diffusion MRI - Part 1: Why low-rank? In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, page 0610, 2013. 142

[158] R. Otazo, E. J. Candès, and D. K. Sodickson. Low-rank and sparse matrix decomposition for accelerated DCE-MRI with background and contrast separation. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM) Workshop on Data Sampling and Image Reconstruction*, Sedona, 2013. 142

[159] J. D. Trzasko and A. Manduca. A unified tensor regression framework for calibrationless dynamic, multi-channel MRI reconstruction. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM)*, volume 21, page 4371, 2013. 146, 149, 154, 157

[160] M. Hutchinson and U. Raff. Fast MRI data acquisition using multiple detectors. *Magnetic Resonance in Medicine*, 6(1):87–91, 1988. 146

[161] J. B. Ra and C. Y. Rim. Fast imaging using subencoding data sets from multiple detectors. *Magnetic Resonance in Medicine*, 30(1):142–5, July 1993. 146

[162] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603, October 1997. 146

[163] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–10, July 2002. 146

[164] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–62, November 1999. 146, 148

[165] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion

MRI. *Magnetic Resonance in Medicine*, 64(3):767–76, September 2010. 148, 153, 157

[166] D. Liang, B. Liu, J. Wang, and L. Ying. Accelerating SENSE using compressed sensing. *Magnetic Resonance in Medicine*, 62(6):1574–84, December 2009. 157

[167] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu. A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging. *Medical Image Analysis*, 15(2):185–201, April 2011. 157

[168] P. J. Shin, P. E. Z. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magnetic Resonance in Medicine*, November 2013. 157

[169] S. G. Lingala, E. DiBella, G. Adluru, C. McGann, and M. Jacob. Accelerating free breathing myocardial perfusion MRI using multi coil radial k-t SLR. *Physics in Medicine and Biology*, 58(20):7309–27, October 2013. 157

[170] L. Wissmann, C. Santelli, W. P. Segars, and S. Kozerke. MRXCAT: Realistic numerical phantoms for cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*, 16(1):63, 2014. 161

[171] J. Lee, Y. Sun, and M. A. Saunders. Proximal Newton-type methods for convex optimization. In *Advances in Neural Information Processing Systems (NIPS)*, pages 827–835, 2012. 162

[172] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, March 1956. 162

[173] S. Ramani, T. Blu, and M. Unser. Monte-Carlo Sure: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Transactions on Image Processing*, 17(9):1540–54, September 2008. 162

[174] D. Goldfarb and Z. Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014. 163

[175] D. G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1969. 165